GENETIC AND PHYSIOLOGICAL REGULATION OF SPECIALIZED METABOLISM AND BIOCHEMICAL DEFENSE IN *ZEA MAYS*

A Dissertation Presented to the Faculty of the Graduate School
of Cornell University
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Shaoqun Zhou
August 2018

GENETIC AND PHYSIOLOGICAL REGULATION OF SPECIALIZED METABOLISM AND BIOCHEMICAL DEFENSE IN *ZEA MAYS*

Shaoqun Zhou, Ph. D.
Cornell University

Maize (*Zea mays*) is one of the most important staple crops worldwide. Every year, significant proportion of the commercial maize production is lost to insect herbivores and phytopathogens despite extensive control measures. This global issue can be alleviated by harnessing the innate biochemical defense mechanisms of maize, which may have been sacrificed over the course of crop domestication for higher yield. In this thesis, I use *Fusarium graminearum*, a widespread fungal pathogen of maize, as a model pathosystem, to study the genetic and physiological control of maize specialized metabolism and biochemical defense. By integrating untargeted metabolomics and transcriptomics data into quantitative genetics framework, I am able to discover novel regulatory genes and mechanisms of specific metabolites, as well as to establish a metabolome-scale resource of metabolite-genetic loci associations at high resolution.

Through comparative metabolomics analyses, I identify two *F. graminearum*-resistance-related acetylated diferuloylsucrose compounds, smiglaside C and smilaside A, which have not been confirmed in maize previously. In an *in vitro* fungal growth inhibition assay, only the di-acetylated smilaside A demonstrates significant bioactivity, whereas the tri-acetylated smiglaside C does not. Genetic mapping of these two compounds, alongside with mutant analyses and physiological experiments, show that ethylene signaling can regulate the metabolism of these two compounds. While ethylene production is required for the accumulation of both compounds *in planta*, their relative abundance is fine-tuned by ethylene sensitivity. Interestingly, the relative abundance, rather than the absolute amount of these two compounds appears to have a more significant influence on maize resistance against *F. graminearum* infection.

In the same genetic mapping population, genetic mapping and metabolite-transcript correlation analyses suggest that a putative vesicular transport protein is a negative regulator of

accumulation of benzaoxazinoids, the most abundant class of specialized metabolites in maize seedlings. This hypothesis is partially supported by genetic mutant analyses and pharmacological disruption of the vesicular transport system *in planta*. However, further experimental evidence is required to establish a role for the vesicular transport system in benzoxazinoid metabolism.

Finally, the chemical genetics approach is extended to a much more diverse maize genome-wide association mapping diversity panel. Multivariate statistical analyses of the large untargeted metabolomics dataset reveal that different classes of specialized metabolites are selectively differentiated between developmental stages and genetic subpopulations. Using liquid chromatography retention time as a proxy of metabolite structure relatedness, it is shown that structurally similar metabolites tend to be co-regulated by shared genetic loci. To demonstrate the utility of the thousands of metabolite-genetic loci association, I experimentally validate that different alleles of a maize citrate synthase gene is responsible for the different structural isomers of hydroxycinnamic acid-hydroxycitric acid conjugates accumulated in the tropical versus temperate maize inbred lines

In summary, work presented in this thesis demonstrates the power of integrating multi-omics dataset to dissect specialized metabolism in maize. It provides examples of metabolic and regulatory gene discovery using a forward genetics approach, and set up a platform for future validation of candidate genetic loci associated with potential metabolites of interest.

BIOGRAPHICAL SKETCH

Shaoqun ‘Simon’ Zhou received his Bachelor of Science degree from University of Washington in Biology in 2012 before moving to Cornell University to pursue a Doctor of Philosophy degree in the Plant Biology Graduate Field. Early in his education, Simon developed a strong interest in traditional Chinese medicine due to the influence from his mother. Through later scientific training in China and the U.S., this early inspiration has transformed into a broader enthusiasm in chemistry-mediated biotic interactions, and has morphed his pantheistic view of Mother Nature. At Cornell University, Simon designed his thesis project on maize specialized metabolism and biochemical defense under the advising of Dr. Georg Jander, Dr. Rebecca Nelson, and Dr. Gary Bergstrom. Outside of laboratories, Simon is an amateur pianist and singer particularly interested in Chinese and Japanese pop music.

Dedicated to my parents

ACKNOWLEDGMENTS

I would like to extend my deepest gratitude to my thesis committee chair Dr. Georg Jander, who has been very open and supportive along the development of my research project. His experiences and creative thinking has been indispensible for the work presented here. I also would like to recognize the persistent support and feedbacks from my thesis committee members, Dr. Rebecca Nelson, and Dr. Gary Bergstrom.

One of the most critical lessons I have learned through working on this project is the importance of cross-disciplinary collaborations. I have drawn experiences from seasoned post-doctoral researchers including Dr. Melkamu Woldermariam, Dr. Jian Yan, Dr. Vered Tzin, and Dr. Annett Richter, as well as talented fellow graduate students such as Dr. Karl Kremling and Ying Kitty Zhang. Learning from these peers has been truly a privilege, and the most enjoyable part of the graduate training process. I am also deeply grateful to Dr. Meena Haribal, Dr. Navid Mohaved, and Ms. Jamie Cummings for patiently sharing their expertise in analytical chemistry and fungus culturing.

My thesis project has been financially supported by the National Science Foundation and a Northeast Sustainable Agriculture Research and Education graduate student grant, and I have been primarily funded by the Teaching Assistantship program at Cornell University. I enjoyed the extensive mentoring experiences with many Cornell undergraduate students and local high school students. Their dedication and assistance have been important contributions to realize my sometimes crazy ideas.

I want to dedicate this thesis to my mother, who has been the original spark that inspired me to pursue a career in biological science, and continue to be my most enthusiastic advocate. The motivation and love in science I derive from her has always been the most important part of my self-depiction. I also want to dedicate this thesis to my late father, who has been the unsung hero that supported my expensive education in the U.S. While our communications have never been as expressive, I know that he has been, and always will be there for me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

*Supplemental datasets provided in a separate electronic file.

LIST OF ABBREVIATIONS

1-ACC	1-aminocyclopropane-1-carboxylate
1-MCP	1-methylcyclopropene
2-ODD	2-oxoglutarate-dependent dioxygenase
ACS	1-aminocyclopropane-1-carboxylate synthase
ANOVA	Analysis of variance
BFA	Brefeldin A
CYP450	Cytochrome P450-dependent monooxygenase
DIBOA	2,4-dihydroxy-1,4-benzoxazin-3-one
DIBOA-Glc	2,4-dihydroxy-1,4-benzoxazin-3-one-β-d-glucopyranose
DIM2BOA	8-*O*-methylated 2-(2,4-dihydroxy-7,8- dimethoxy-1,4-benzoxazin-3-one)
DIM2BOA-Glc	8-*O*-methylated 2-(2,4-dihydroxy-7,8- dimethoxy-1,4-benzoxazin-3-one)-β-d-glucopyranose
DIMBOA	2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one
DIMBOA-Glc	2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one-β-d-glucopyranose
DMSO	Dimethylsulfoxide
EIN2	Ethylene Insensitive 2
ELISA	Enzyme-linked immuno-sorbent assay
ER	Endoplasmic reticulum
ESI	Electron spray ionization
FDR	False discovery rate
GA	Gibberellic acid
GFP	Green fluorescent protein
GWAS	Genome-wide association study
HDM2BOA-Glc	2-(2-hydroxy-4,7,8- trimethoxy-1,4-benzoxazin-3-one)-β-d-glucopyranose
HDMBOA	2-(2-hydroxy-4,7-dimethoxy-1,4-benzoxazin-3-one)
HDMBOA-Glc	2-(2-hydroxy-4,7-dimethoxy-1,4-benzoxazin-3-one)-β-d-glucopyranose
HMBOA	Hydroxy-7-methoxy-1,4-benzoxazin-3-one
IBM	Intermated B73 x Mo17
Kbps	Kilo base pairs
LC	Liquid chromatography
LD	Linkage disequilibrium
LOD	Log of odds
m/z	Mass-to-charge ratio
MBOA	2-methyl-3,1-benzoxazin-4-one
Mbps	Mega base pairs
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
N.D.	Not detected
N.S.	Not significant
NAM	Nested association mapping
NIL	Near-isogenic line
NMR	Nuclear magnetic resonance

OMT	*O*-methyltransferase
PCA	Principal component analysis
qRT-PCR	Quantitative reverse transcriptase polymerase chain reaction
QTL	Quantiative trait loci
RIL	Recombinant inbred line
S.d.	Standard deviation
S.e.	Standard error
SNP	Single nucleotide polymorphism
TUB	Tubulin
UFMu	Uniform-*Mutator*
UTR	Un-translated region
UV	Ultraviolet

CHAPTER ONE

SPECIALIZED METABOLISM AND BIOCHEMICAL DEFENSE IN MAIZE: AN OVERVIEW

Plants in natural and manmade ecosystems are constantly exposed to herbivorous animals and phytopathogenic microbes. As sessile organisms that cannot escape from their natural enemies, plants cope with this challenge by evolving specialized metabolic pathways that give rise to a diverse arsenal of defensive and deterrent compounds, such as nicotine in tobacco (*Nicotiana tabacum*) and glucosinolates in *Arabidopsis thaliana* (Baldwin et al. 1998; Mewis et al. 2005). In addition to their ecological functions, plant-derived specialized metabolites are also important source of essential nutrients and pharmacological products for animals and human society, such as anthocyanins, carotenoids, and salicylic acids (Butelli et al. 2008; Cazzonelli 2011; Vlot et al. 2009). Therefore, understanding and subsequent engineering of plant specialized metabolism are of great scientific and economic interest.

For my Ph.D. thesis, I have used maize (*Zea mays*) as a model system. In North America, maize is the most important staple crop with over 13 billion bushels produced per annum, of which approximately 10% is lost to various diseases (Mueller 2016a, b, 2017). Maize is also known for its great genetic diversity, involving both nucleotide polymorphisms and structural genomic variation (Buckler et al. 2006; Jiao et al. 2017). The genetic architecture of disease resistance in maize has been investigated extensively using publicly available genetic resources (Benson et al. 2015; Mideros et al. 2012; Olukolu et al. 2014). The significance of specialized metabolism in industrial maize production is exemplified by the association of specific biosynthetic genes with resistance against insect pests and phytopathogens (Handrick et al. 2016; Meihls et al. 2013; Yang et al. 2017b). Such studies highlight the potential of enlisting naturally occurring specialized metabolites in crop species to enhance quantitative disease resistance. Compared to foliar and ear diseases, maize seedling diseases remain a relatively understudied area, even though in some years they can account for more yield loss than any single disease in the aboveground tissues (Mueller 2016a). This may be due to the fact that experimental methods developed for large scale screening of diseases in aboveground tissues, such as controlled pathogen inoculation and visual symptom scoring, are difficult to apply to seedling diseases under field conditions. Furthermore, like in many other plant species, young seedling tissues of maize contain the highest amount of specialized metabolites, probably because the seedlings

cannot afford extensive tissue loss from pathogen infection and the physical barriers against pathogen establishment are relatively weak at this early developmental stage (Barton and Boege 2017). Therefore, maize-pathogen interaction at the seedling stage provides an attractive and economically important system to study the genetic and physiological regulation of specialized metabolism.

Benzoxainoids

Perhaps the most well studied specialized metabolites in maize are the benzoxazinoids. These indole-derived small molecules are primarily found in maize, wheat (*Triticum spp.*), rye (*Secale cereal*), and other wild and cultivated Poaceae species (Barria et al. 1992; Copaja et al. 1991a; Copaja et al. 1991b; Gianoli and Niemeyer 1998a, b; Ni and Quisenberry 2000; Nicol et al. 1992; Niemeyer 1988; Niemeyer et al. 1992; Singh et al. 2003; Wu et al. 2001). This is a family of approximately 20 compounds share the 2-hydroxy-2*H*-1,4-benzoxazin-3(4*H*)-one skeleton, and the dominant species could account for up to 0.3% of fresh weight in some tissues (Meihls et al. 2013).

Efforts to elucidate the biosynthetic pathway of benzoxazinoids date back to as early as the 1960s. After 40 years of active research, the entire pathway leading to the biosynthesis of 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA), the prevalent benzoxazinoid compound in most of the crop species, has been fully characterized (Niemeyer 2009). In brief, the first committed step of benzoxazinoid biosynthesis is believed to be the production of indole from indole-3-glycerolphosphate by the enzyme *BX1*. The indole molecules are oxidized by four subsequent cytochrome P450-dependent monooxygenases (CYP450, *BX2-5*) in the endoplasmic reticulum, producing the simplest benzoxazinoid, 2,4-dihydroxy-1,4-benzoxazin-3-one (DIBOA). Two redundant UDP-glucosyltransferases, *BX8 and BX9,* then attach a glucose moiety to the cytotoxic DIBOA to produce the less toxic glucoside. 2,4-dihydroxy-1,4-benzoxazin-3-one-β-d-glucopyranose (DIBOA-Glc) is further oxidized by a 2-oxoglutarate-dependent dioxygenase (2-ODD, *BX6*) and methylated by an *O*-methyltransferase (OMT, *BX7*) to produce DIMBOA-glucoside.

More recently, the quantitative genetics approach has been successfully adopted to characterize the later biosynthetic steps of the more decorated benzoxazinoids. By genetic mapping with 2-(2-hydroxy-4,7-dimethoxy-1,4-benzoxazin-3-one)-β-d-glucopyranose

(HDMBOA-Glc), Meihls et al. (2013) have identified another OMT, *BX10*, and its two paralogs *BX11* and *BX12*. Taking the same experimental approach, Handrick et al. (2016) have identified *BX13*, a *BX6*-like 2-ODD, and *BX14*, a *BX10*-like OMT, that are required to produce 8-*O*-methylated 2-(2,4-dihydroxy-7,8-dimethoxy-1,4-benzoxazin-3-one)-β-d-glucopyranose (DIM_2BOA-Glc) and 2-(2-hydroxy-4,7,8-trimethoxy-1,4-benzoxazin-3-one)-β-d-glucopyranose (HDM_2BOA-Glc). Furthermore, Wisecaver et al. (2017) recently found that a putative indole-3-glycerolphosphate synthase gene, unlike its functional homologs in maize, significantly co-express with other benzoxazinoid biosynthetic genes. This observation leads to the hypothesis that the committing step in benzoxazinoid biosynthesis may occur prior to indole production by *BX1*. All benzoxazinoid glucosides are stored in vacuoles. Upon tissue disruption, these glucosides are exposed to specific glucosidases in the plastids, which cleave off the glucosyl groups, producing the bioactive aglucones.

Benzoxazinoids are developmentally regulated. In dry wheat and maize seeds, benzoxazinoids are not detected, but only found after seedling germination (Cambier et al. 2000; Nakagawa et al. 1995; Villagrasa et al. 2006). Levels of these compounds peak within the first two weeks after germination, and steadily decline as plants age. Recently, Zheng et al. (2015) identified a *cis* regulatory element approximately 140 Kbps upstream of the *ZmBX1* that significantly prolongs its expression in maize seedlings. Interestingly, this distal *cis* element specifically regulates the age-dependent expression of *ZmBX1*, but not *ZmBX2*, which is located only 2.5 Kbps downstream. On top of their temporal dynamics, benzoxazinoid compounds are also non-uniformly distributed in different plant organs. Abundance of benzoxazinoid compounds in maize seedling shoots and roots vary greatly across different maize genotypes, and there is no clear correlation in the benzoxazinoid abundance between these two organs (Handrick et al. 2016). At a finer tissue level, Ahmad et al. (2011) reported that DIMBOA aglucone was primarily found in leaf apoplast, though the glucosides were usually found at higher levels in whole tissues.

As a class of defense-related specialized metabolites, benzoxazinoids are typically considered as phytoanticipins, which are constitutively produced by plants even in absence of an imminent threat. Yet, transcriptomics and metabolomics studies have revealed active dynamics of benzoxazinoid metabolism in face of biotic stressors or their plant-derived signals. Both chewing caterpillars and phloem-sucking aphids can induce accumulation of benzoxazinoid

compounds and upregulation of their biosynthetic genes in maize (Tzin et al. 2015; Tzin et al. 2017). Similarly, benzoxazinoid compounds can also be induced by various fungal pathogens in wheat and maize, and the magnitude of induction is a function of host resistance, pathogen virulence, and possibly environmental conditions (Oikawa et al. 2004; Weibull and Niemeyer 1995). In addition to inducing overall benzoxazinoid accumulation, caterpillar feeding and fungal pathogen infection can induce conversion of DIMBOA-Glc into its methylated derivative HDMBOA-Glc (Huffaker et al. 2011; Oikawa et al. 2002). This inducible conversion has been proposed to enhance plant resistance as HDMBOA aglucone has stronger bioactivity than DIMBOA. Similar induced changes in benzoxazinoid metabolism can be achieved with a plant elicitor peptide and defense-related phytohormones including jasmonic acid, methyl jasmonate, and *cis*-jasmone (Moraes et al. 2008; Oikawa et al. 2002; Slesak et al. 2001; Wang et al. 2007).

Benzoxazinoid compounds have been shown to be important for maize defense against aphids and a fungal pathogen (Ahmad et al. 2011; Handrick et al. 2016; Meihls et al. 2013). It has also been shown that survival of various strains and species of *Fusarium* fungal pathogens are positively correlated with their capability to degrade bioactive benzoxazinoid compounds (Glenn et al. 2001).

Hydroxycinnamic acids

In addition to the mostly grass-specific benzoxazinoid compounds, maize also produces specialized metabolites that are shared with other plant families. Hydroxycinnamic acids are a group of small organic acids derived from phenylalanine and tyrosine through the phenylpropanoid pathway (Kanehisa et al. 2017; Kanehisa et al. 2016). The first step of their biosynthesis is the removal of the amine group by phenylalanine ammonia lyase to produce cinnamic acid, or the same reaction on tyrosine by a bi-functional phenylalanine/tyrosine ammonia lyase to produce *p*-coumaric acid. These hydroxycinnamic acids are linked to co-enzyme A through promiscuous coumarate-CoA ligases, and can then be conjugated to various small molecules including quinic acid and shikimic acid through a diverse family of hydroxycinnamoyl transferases. In other plant species, *p*-coumaric acid and its CoA conjugate can be directly hydroxylated to produce caffeic acid, and subsequently ferulic acid, hydroxyferulic acid, and sinapic acid and their respective CoA conjugates through a series of methylation and hydroxylation on the phenol ring. Finally, these hydroxycinnamic acids are also

important cross-linking agents for cellulose fibers, and are the chemical precursors to various lignin monomers (Ralph 2010). Through these means, they are significant contributors to the fortification of plant physical defense.

Regulation of the phenylpropanoid pathway at the molecular level has been an active area of research. In maize, two MYB-family transcription factors, *ZmMYB31* and *ZmMYB42,* have been shown to suppress the activity of a caffeoyl-*O*-methyltransferase (Fornale et al. 2010; Sonbol et al. 2009). Similar repressive regulation of phenylpropanoid pathway genes by MYB transcription factors has been observed throughout the plant kingdom (Liu et al. 2015). Based on these observations, a promoter-transcription factor interaction network approach has recently been taken with the aim of identifying transcriptional regulators of the phenylpropanoid pathway in maize at the global level (Yang et al. 2017a).

Many hydroxycinnamic acids as well as their various conjugates have been identified as plant disease resistance-related compounds, either through comparative studies, or *in vitro* bioassays. For example, caffeoyl-quinic acid (*i.e.* chlorogenic acid), ferulic acid, and diferulic acid contents in maize have been associated with resistance against various fungal pathogens in the *Fusarium* genus (Santiago et al. 2007). Furthermore, ferulic acid has been shown to inhibit *F. graminearum* growth *in vitro* (Ponts et al. 2011). Interestingly, the same study shows that *F. graminearum* in culture responds to ferulic acid addition by increasing mycotoxin production, further suggesting that this compound imposes stress on the phytopathogenic fungus.

Flavonoids

Also derived from the phenylpropanoid pathway is another large and diverse family of metabolites called the flavonoids. These compounds share a 15-carbon backbone with two phenolic rings and a heterocyclic ring. The biosynthetic pathway of major flavonoids in maize has been well characterized. It starts by combining *p*-coumaroyl-CoA and malonyl-CoA with chalcone synthase, and formation of the heterocyclic ring with chalcone isomerase to give rise to the simplest flavonoid, naringenin (Kanehisa et al. 2017; Kanehisa et al. 2016). Various hydroxylation and methylation can then take place on different carbons to generate diverse flavonoid compounds. Interestingly, bioactive flavonoids often exist as glycosides *in vivo*, and the number, type, and the position of the sugar moiety can have profound impact on the bioactivity of the molecule (Xiao 2017). More recently, a couple of metabolome-scale genetic

mapping studies have provided a large number of novel candidate genes involved in the biosynthesis and regulation of flavonoids in maize (Jin et al. 2017; Wen et al. 2014).

Bioactivity of maize flavonoids against biotic stress specifically is not as well documented compared to benzoxazinoids and hydroxycinnamic acids. One of the best-characterized examples is the anti-insect function of maysin (Gueldner et al. 1991). Indirect evidences of potential roles of flavonoids in disease resistance primarily come from pathogen induction assays. However, in a two-years field study with maize isogenic lines of differing flavonoid pigmentation in the ears, the authors concluded that flavonoids by themselves are incapable of enhancing resistance against *Fusarium* pathogens and mycotoxin accumulation (Venturini et al. 2015).

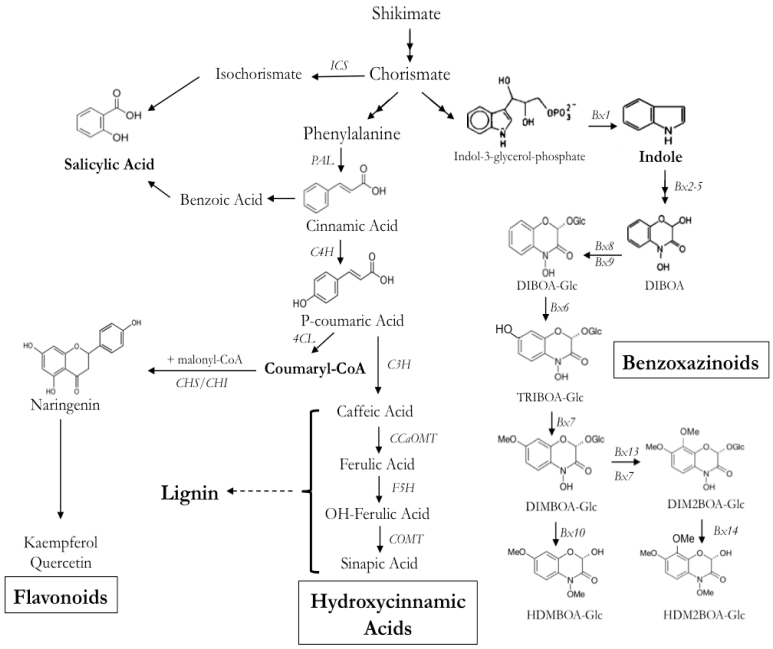


Figure 1.1 Major classes of maize specialized metabolites originated from the shikimate/chorismate pathway.

All three classes of specialized metabolites described above can be traced back to the shikimate/chorismate pathway (Figure 1.1). They can be readily extracted from maize tissues with mid-polarity solvents (>50% methanol), separated by reverse phase chromatography, and detected with either electron spray ionization mass spectrometry or photodiode array detectors. Furthermore, each class can be identified with its characteristic UV absorbance profile, though the specific metabolites within each class are hard to distinguish through this mean. All of the untargeted metabolomics analyses presented in this thesis are initially modified from a high-performance liquid chromatography protocol optimized for benzoxazinoids detection and quantification based on UV absorbance. Hence, I am well aware that important and diverse specialized metabolites such as the terpenoids are largely omitted from my analyses due to limitation in extraction and analytical methods. Additionally, I also recognize that there are other less studied specialized metabolites in maize

such as phenolamides that are probably included in my metabolite extract, but haven't been the focus of my subsequent analyses.

In the following chapters of this thesis, I present four related studies investigating the specialized metabolism and/or biochemical defense of maize using comparative morphology, metabolomics, transcriptomics, and quantitative genetics methods. In the first two chapters, I use maize seedling root-*Fusarium graminearum* as a model patho-system. Specifically, I establish a series of morphological markers indicative of natural variation in *F. graminearum*-resistance level by surveying diverse maize genotypes and *F. graminearum* isolates (Chapter Two). I use one of these markers to identify *F. graminearum*-resistant and susceptible maize inbred lines with existing genetic resources, performed comparative metabolomics analyses to identify novel *F. graminearum*-resistance-related maize specialized metabolites, and investigate their genetic and physiological regulation with a quantitative genetics approach (Chapter Three). I further re-analyze the untargeted metabolomics and transcriptomics dataset generated for this genetic mapping population on a global level, identifying genomic hotspots associated multiple metabolic traits, and validating one of them through genetic mutants analyses and pharmacological treatment experiments (Chapter Four). Finally, I extend the same metabolome-scale genetic mapping and cross-omics correlation analyses approach to a much more diverse maize genome-wide association study panel to explore genetic control of maize specialized metabolism at the metabolome scale (Chapter Five). Throughout these studies, I have been constantly challenged to pick up data analytics skills and apply my creativity to approach the massive datasets with the right questions and hypotheses. Most importantly, I have been trained to think quantitatively, which allowed me to see the limits in current qualitative models in plant physiology and molecular biology. These opinions are summarized in Chapter Six.

Reference

Ahmad S, Veyrat N, Gordon-Weeks R, Zhang YH, Martin J, Smart L, Glauser G, Erb M, Flors V, Frey M, Ton J (2011) Benzoxazinoid metabolites regulate innate immunity against aphids and fungi in maize. Plant Physiol 157 (1):317-327. doi:10.1104/pp.111.180224

Baldwin IT, Gorham D, Schmelz EA, Lewandowski CA, Lynds GY (1998) Allocation of nitrogen to an inducible defense and seed production in *Nicotiana attenuata*. Oecologia 115 (4):541-552. doi:DOI 10.1007/s004420050552

Barria BN, Copaja SV, Niemeyer HM (1992) Occurrence of DIBOA in wild Hordeum species and its relation to aphid resistance. Phytochemistry 31 (1):89-91. doi:Doi 10.1016/0031-9422(91)83012-A

Barton KE, Boege K (2017) Future directions in the ontogeny of plant defence: understanding the evolutionary causes and consequences. Ecol Lett 20 (4):403-411. doi:10.1111/ele.12744

Benson JM, Poland JA, Benson BM, Stromberg EL, Nelson RJ (2015) Resistance to gray leaf spot of maize: genetic architecture and mechanisms elucidated through nested association mapping and near-isogenic line analysis. PLoS Genet 11 (3). doi:ARTN e100504510.1371/journal.pgen.1005045

Buckler ES, Gaut BS, McMullen MD (2006) Molecular and functional diversity of maize. Curr Opin Plant Biol 9 (2):172-176. doi:10.1016/j.pbi.2006.01.013

Butelli E, Titta L, Giorgio M, Mock HP, Matros A, Peterek S, Schijlen EGWM, Hall RD, Bovy AG, Luo J, Martin C (2008) Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nat Biotechnol 26 (11):1301-1308. doi:10.1038/nbt.1506

Cambier V, Hance T, de Hoffmann E (2000) Variation of DIMBOA and related compounds content in relation to the age and plant organ in maize. Phytochemistry 53 (2):223-229. doi:Doi 10.1016/S0031-9422(99)00498-7

Cazzonelli CI (2011) Carotenoids in nature: insights from plants and beyond. Funct Plant Biol 38 (11):833-847. doi:10.1071/Fp11192

Copaja SV, Barria BN, Niemeyer HM (1991a) Hydroxamic acid content of perennial Triticeae. Phytochemistry 30 (5):1531-1534. doi:Doi 10.1016/0031-9422(91)84202-4

Copaja SV, Niemeyer HM, Wratten SD (1991b) Hydroxamic acid levels in Chilean and British wheat seedlings. Ann Appl Biol 118 (1):223-227. doi:DOI 10.1111/j.1744-7348.1991.tb06100.x

Fornale S, Shi X, Chai C, Encina A, Irar S, Capellades M, Fuguet E, Torres JL, Rovira P, Puigdomenech P, Rigau J, Grotewold E, Gray J, Caparros-Ruiz D (2010) ZmMYB31 directly represses maize lignin genes and redirects the phenylpropanoid metabolic flux. Plant J 64 (4):633-644. doi:10.1111/j.1365-313X.2010.04363.x

Gianoli E, Niemeyer HM (1998a) Allocation of herbivory-induced hydroxamic acids in the wild wheat Triticum uniaristatum. Chemoecology 8 (1):19-23. doi:Doi 10.1007/Pl00001799

Gianoli E, Niemeyer HM (1998b) DIBOA in wild Poaceae: Sources of resistance to the Russian wheat aphid (*Diuraphis noxia*) and the greenbug (*Schizaphis graminum*). Euphytica 102 (3):317-321. doi:Doi 10.1023/A:1018323121213

Glenn AE, Hinton DM, Yates IE, Bacon CW (2001) Detoxification of corn antimicrobial compounds as the basis for isolating Fusarium verticillioides and some other Fusarium species from corn. Appl Environ Microbiol 67 (7):2973-2981. doi:10.1128/AEM.67.7.2973-2981.2001

Gueldner RC, Snook ME, Wiseman BR, Widstrom NW, Himmelsbach DS, Costello CE (1991) Maysin in Corn, Teosinte, and Centipede Grass. Acs Sym Ser 449:251-263

Handrick V, Robert CAM, Ahern KR, Zhou SQ, Machado RAR, Maag D, Glauser G, Fernandez-Penny FE, Chandran JN, Rodgers-Melnik E, Schneider B, Buckler ES, Boland W, Gershenzon J, Jander G, Erb M, Kollner TG (2016) Biosynthesis of 8-*O*-methylated benzoxazinoid defense compounds in maize. Plant Cell 28 (7):1682-1700. doi:10.1105/tpc.16.00065

Huffaker A, Kaplan F, Vaughan MM, Dafoe NJ, Ni XZ, Rocca JR, Alborn HT, Teal PEA, Schmelz EA (2011) Novel acidic sesquiterpenoids constitute a dominant class of pathogen-induced phytoalexins in maize. Plant Physiol 156 (4):2082-2097. doi:10.1104/pp.111.179457

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS, Guill K, Regulski M, Kumari S, Olson A, Gent J, Schneider KL, Wolfgruber TK, May MR, Springer NM, Antoniou E, McCombie WR, Presting GG, McMullen M, Ross-Ibarra J, Dawe RK, Hastie A, Rank DR, Ware D (2017) Improved maize reference genome with single-molecule technologies. Nature 546 (7659):524-527. doi:10.1038/nature22971

Jin M, Zhang XH, Zhao MC, Deng M, Du YH, Zhou Y, Wang SC, Tohge T, Fernie AR, Willmitzer L, Brotman Y, Yan JB, Wen WW (2017) Integrated genomics-based mapping reveals the genetics underlying maize flavonoid biosynthesis. Bmc Plant Biol 17. doi:ARTN 1710.1186/s12870-017-0972-z

Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45 (D1):D353-D361. doi:10.1093/nar/gkw1092

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44 (D1):D457-D462. doi:10.1093/nar/gkv1070

Liu JY, Osbourn A, Ma PD (2015) MYB Transcription factors as regulators of phenylpropanoid metabolism in plants. Mol Plant 8 (5):689-708. doi:10.1016/j.molp.2015.03.012

Meihls LN, Handrick V, Glauser G, Barbier H, Kaur H, Haribal MM, Lipka AE, Gershenzon J, Buckler ES, Erb M, Kollner TG, Jander G (2013) Natural variation in maize aphid resistance is associated with 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one glucoside methyltransferase activity. Plant Cell 25 (6):2341-2355. doi:10.1105/tpc.113.112409

Mewis I, Appel HM, Hom A, Raina R, Schultz JC (2005) Major signaling pathways modulate Arabidopsis glucosinolate accumulation and response to both phloem-feeding and chewing insects. Plant Physiol 138 (2):1149-1162. doi:10.1104/pp.104.053389

Mideros SX, Windham GL, Williams WP, Nelson RJ (2012) Tissue-specific components of resistance to Aspergillus ear rot of Maize. Phytopathology 102 (8):787-793. doi:10.1094/Phyto-12-11-0355

Moraes MCB, Birkett MA, Gordon-Weeks R, Smart LE, Martin JL, Pye BJ, Bromilow R, Pickett JA (2008) cis-Jasmone induces accumulation of defence compounds in wheat, *Triticum aestivum*. Phytochemistry 69 (1):9-17. doi:10.1016/j.phytochem.2007.06.020

Mueller D (2016a) Corn disease loss estimates from the United States and Ontario, Canada - 2014. Purdue University Extension, Retrieved from http://cropprotectionnetwork.org/crop-loss-estimates/corn-disease-loss-estimates-2014/.

Mueller D (2016b) Corn disease loss estimates from the United States and Ontario, Canada - 2015. Purdue University Extension, Retrieved from http://cropprotectionnetwork.org/crop-loss-estimates/corn-disease-loss-estimates-2015/.

Mueller D (2017) Corn disease loss estimates from the United States and Ontario, Canada - 2016. Purdue University Extension, Retrieved from http://cropprotectionnetwork.org/crop-loss-estimates/corn-disease-loss-estimates-2016/.

Nakagawa E, Amano T, Hirai N, Iwamura H (1995) Noninduced cyclic hydroxamic acids in wheat during juvenile stage of growth. Phytochemistry 38 (6):1349-1354. doi:Doi 10.1016/0031-9422(94)00831-D

Ni XZ, Quisenberry SS (2000) Comparison of DIMBOA concentrations among wheat isolines and corresponding plant introduction lines. Entomol Exp Appl 96 (3):275-279. doi:DOI 10.1046/j.1570-7458.2000.00706.x

Nicol D, Copaja SV, Wratten SD, Niemeyer HM (1992) A screen of worldwide wheat cultivars for hydroxamic acid levels and aphid antixenosis. Ann Appl Biol 121 (1):11-18. doi:DOI 10.1111/j.1744-7348.1992.tb03982.x

Niemeyer HM (1988) Hydroxamic acid content of Triticum species. Euphytica 37 (3):289-293

Niemeyer HM (2009) Hydroxamic acids derived from 2-Hydroxy-2H-1,4-Benzoxazin-3(4H)-one: key defense chemicals of cereals. J Agr Food Chem 57 (5):1677-1696. doi:10.1021/jf8034034

Niemeyer HM, Copaja SV, Barria BN (1992) The Triticeae as sources of hydroxamic acids, secondary metabolites in wheat conferring resistance against aphids. Hereditas 116 (3):295-299. doi:DOI 10.1111/j.1601-5223.1992.tb00158.x

Oikawa A, Ishihara A, Iwamura H (2002) Induction of HDMBOA-Glc accumulation and DIMBOA-Glc 4-O-methyltransferase by jasmonic acid in poaceous plants. Phytochemistry 61 (3):331-337. doi:Pii S0031-9422(02)00225-X

Oikawa A, Ishihara A, Tanaka C, Mori N, Tsuda M, Iwamura H (2004) Accumulation of HDMBOA-Glc is induced by biotic stresses prior to the release of MBOA in maize leaves. Phytochemistry 65 (22):2995-3001. doi:10.1016/j.phytochem.2004.09.006

Olukolu BA, Wang GF, Vontimitta V, Venkata BP, Marla S, Ji J, Gachomo E, Chu K, Negeri A, Benson J, Nelson R, Bradbury P, Nielsen D, Holland JB, Balint-Kurti PJ, Johal G (2014) A genome-wide association study of the maize hypersensitive defense response identifies genes that cluster in related pathways. PLoS Genet 10 (8):e1004562. doi:10.1371/journal.pgen.1004562

Ponts N, Pinson-Gadais L, Boutigny AL, Barreau C, Richard-Forget F (2011) Cinnamic-derived acids significantly affect *Fusarium graminearum* growth and *in vitro* synthesis of Type B trichothecenes. Phytopathology 101 (8):929-934. doi:10.1094/Phyto-09-10-0230

Ralph J (2010) Hydroxycinnamates in lignification. Phytochem Rev 9 (1):65-83. doi:10.1007/s11101-009-9141-9

Santiago R, Reid LM, Arnason JT, Zhu XY, Martinez N, Malvar RA (2007) Phenolics in maize genotypes differing in susceptibility to Gibberella stalk rot (*Fusarium graminearum* Schwabe). J Agr Food Chem 55 (13):5186-5193. doi:10.1021/jf070641e

Singh P, Suman A, Shrivastava AK (2003) Isolation and identification of allelochemicals from sugarcane leaves. Allelopathy J 12 (1):71-79

Slesak E, Slesak M, Gabrys B (2001) Effect of methyl jasmonate on hydroxamic acid content, protease activity, and bird cherry-oat aphid *Rhopalosiphum padi* (L.) probing behavior. J Chem Ecol 27 (12):2529-2543. doi:Doi 10.1023/A:1013635717049

Sonbol FM, Fornale S, Capellades M, Encina A, Tourino S, Torres JL, Rovira P, Ruel K, Puigdomenech P, Rigau J, Caparros-Ruiz D (2009) The maize ZmMYB42 represses the phenylpropanoid pathway and affects the cell wall structure, composition and degradability in *Arabidopsis thaliana*. Plant Mol Biol 70 (3):283-296. doi:10.1007/s11103-009-9473-2

Tzin V, Fernandez-Pozo N, Richter A, Schmelz EA, Schoettner M, Schafer M, Ahern KR, Meihls LN, Kaur H, Huffaker A, Mori N, Degenhardt J, Mueller LA, Jander G (2015) Dynamic maize responses to aphid feeding are revealed by a time series of transcriptomic and metabolomic assays. Plant Physiol 169 (3):1727-1743. doi:10.1104/pp.15.01039

Tzin V, Hojo Y, Strickler SR, Bartsch LJ, Archer CM, Ahern KR, Zhou SQ, Christensen SA, Galis I, Mueller LA, Jander G (2017) Rapid defense responses in maize leaves induced by Spodoptera exigua caterpillar feeding. J Exp Bot 68 (16):4709-4723. doi:10.1093/jxb/erx274

Venturini G, Toffolatti SL, Assante G, Babazadeh L, Campia P, Fasoli E, Salomoni D, Vercesi A (2015) The influence of flavonoids in maize pericarp on fusarium ear rot symptoms and fumonisin accumulation under field conditions. Plant Pathol 64 (3):671-679. doi:10.1111/ppa.12304

Villagrasa M, Guillamon M, Labandeira A, Taberner A, Eljarrat E, Barcelo D (2006) Benzoxazinoid allelochemicals in wheat: Distribution among foliage, roots, and seeds. J Agr Food Chem 54 (4):1009-1015. doi:10.1021/jf050898h

Vlot AC, Dempsey DA, Klessig DF (2009) Salicylic acid, a multifaceted hormone to combat disease. Annu Rev Phytopathol 47:177-206. doi:10.1146/annurev.phyto.050908.135202

Wang JW, Xu T, Zhang LW, Zhong ZM, Luo SM (2007) Effects of methyl jasmonate on hydroxamic acid and phenolic acid content in maize and its allelopathic activity to *Echinochloa crusgalli* (L.). Allelopathy J 19 (1):161-169

Weibull J, Niemeyer HM (1995) Changes in dihydroxymethoxybenzoxazinone glycoside content in wheat plants infected by 3 plant-pathogenic fungi. Physiol Mol Plant P 47 (3):201-212. doi:DOI 10.1006/pmpp.1995.1052

Wen WW, Li D, Li X, Gao YQ, Li WQ, Li HH, Liu J, Liu HJ, Chen W, Luo J, Yan JB (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nat Commun 5. doi:ARTN 343810.1038/ncomms4438

Wisecaver JH, Borowsky AT, Tzin V, Jander G, Kliebenstein DJ, Rokas A (2017) A global co-expression network approach for connecting genes to specialized metabolic pathways in plants. Plant Cell 29 (5):944-959. doi:10.1105/tpc.17.00009

Wu HW, Haig T, Pratley J, Lemerle D, An M (2001) Allelochemicals in wheat (*Triticum aestivum* L.): Production and exudation of 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one. J Chem Ecol 27 (8):1691-1700. doi:Doi 10.1023/A:1010422727899

Xiao JB (2017) Dietary flavonoid aglycones and their glycosides: Which show better biological significance? Crit Rev Food Sci 57 (9):1874-1905. doi:10.1080/10408398.2015.1032400

Yang F, Li W, Jiang N, Yu HD, Morohashi K, Ouma WZ, Morales-Mantilla DE, Gomez-Cano FA, Mukundi E, Prada-Salcedo LD, Velazquez RA, Valentin J, Mejia-Guerra MK, Gray J, Doseff AI, Grotewold E (2017a) A maize gene regulatory network for phenolic metabolism. Mol Plant 10 (3):498-515. doi:10.1016/j.molp.2016.10.020

Yang Q, He Y, Kabahuma M, Chaya T, Kelly A, Borrego E, Bian Y, El Kasmi F, Yang L, Teixeira P, Kolkman J, Nelson R, Kolomiets M, J LD, Wisser R, Caplan J, Li X, Lauter N, Balint-Kurti P (2017b) A gene encoding maize caffeoyl-CoA O-methyltransferase confers quantitative resistance to multiple pathogens. Nat Genet 49 (9):1364-1372. doi:10.1038/ng.3919

Zheng LL, McMullen MD, Bauer E, Schon CC, Gierl A, Frey M (2015) Prolonged expression of the BX1 signature enzyme is associated with a recombination hotspot in the benzoxazinoid gene cluster in *Zea mays*. J Exp Bot 66 (13):3917-3930. doi:10.1093/jxb/erv192

CHAPTER TWO

FUSARIUM GRAMINEARUM-INDUCED SHOOT ELONGATION AND ROOT REDUCTION IN MAIZE SEEDLINGS CORRELATE WITH LATER SEEDLING BLIGHT SEVERITY[1, 2]

Abstract

Fusarium graminearum seedling blight is a common disease of maize (*Zea mays*). Development of genetic resistance to seedling blight in maize germplasm requires efficient and accurate quantitative assessment of disease severity. Through artificial inoculation experiments under controlled growth conditions, we determined that host genotype, pathogen genotype, and infection dose influence the extent to which *F. graminearum* induces shoot elongation and inhibits root growth in maize seedlings. A comparison of fifteen maize inbred lines showed independent variation of these two fungus-induced effects on seedling growth. In a broader survey with nine commercial maize hybrids and three field-collected fungal isolates, there was significant correlation between these seedling growth responses, as well as with later seedling blight severity. Analysis of variance suggested that this variation and the observed correlative relationships were primarily driven by differing pathogenicity of the three fungal isolates. Together, our results indicate that *F. graminearum*-induced shoot elongation and root reduction in maize seedlings have distinct underlying physiological mechanisms, and that early observations of seedling growth responses can serve as a proxy for investigating natural variation in host resistance and pathogen aggressiveness at later growth stages.

[1] All work in this chapter has been published in *Plant Direct* DOI: 10.1002/pld3.75

[2] Shaoqun Zhou designed and performed all experiments. Justin Bae assisted in root imaging data acquisition and analyses. Dr. Gary Bergstrom and Dr. Georg Jander provided experimental materials and advising and feedback throughout the project

Introduction

Fusarium graminearum Schwabe (teleomorph: *Gibberella zeae*) is a widespread fungal phytopathogen that causes seedling blight, root rot, stem rot, or ear rot in maize (*Zea mays*), depending on the timing and tissue of infection (Munkvold and White 2016). The severity of maize diseases caused by *F. graminearum* in the United States varies significantly from year to year, causing 15 to 80 million bushels of annual yield loss due to ear rot, and another 30 to 85 million bushels due to stalk rot (Mueller 2016a, b, c, 2017). Unlike Gibberella ear rot and stem rot diseases caused by *F. graminearum*, which can be tracked by the characteristic pink-reddish pigmentation of fungal mycelium, seedling blight and root rot diseases are often caused by a consortium of soil-dwelling fungal and oomycete pathogens, including *F. graminearum*. While the dominant species/strains of these consortia are likely to vary considerably across geographic locations, collectively these seedling blight and root rot diseases can account for over 200 million bushels of annual yield loss, making them among the most damaging maize diseases in the northern states (Mueller 2016b). Management of *F. graminearum* seedling blight can be difficult because standard seed fungicide treatment is not always sufficient and may promote the development of resistance in the fungal population. Genetic sources of seedling blight resistance have yet to be established in commercial maize germplasm.

In addition to causing direct yield loss through tissue rotting, *F. graminearum* can also cause crop contamination with mycotoxins such as deoxynivalenol and zearalenone, leaving the grain and stover unsafe for human and livestock consumption, and imposing further financial burdens on the growers (Arunachalam and Doohan 2013; Maresca 2013). Furthermore, *F. graminearum* can also cause destructive scab and head blight diseases in barley and wheat (McMullen et al. 2012; Trail 2009). Considering the overwinter persistence and airborne infectious ascospores of this pathogen, unchecked *F. graminearum* diseases in maize could have a far-reaching negative impact on other crop species in the vicinity.

Breeders and researchers interested in improving resistance against *F. graminearum* in maize have been doing artificial inoculation experiments in both field and controlled environments, primarily focused on infected maize ears and stalks. Genetic mapping of Gibberella ear rot resistance based on symptom severity has identified a large number of quantitative trait loci (QTL), each with small effect size. Moreover, these QTL tend to be sensitive to experimental methods and environmental conditions, making them difficult to

validate across different studies (Ali et al. 2005; Brauner et al. 2017; Kebede et al. 2016). In contrast, major genetic loci contributing to Gibberella stalk rot resistance have been identified with bi-parental mapping populations (Chen et al. 2017; Ma et al. 2017; Yang et al. 2010; Zhang et al. 2012). Ye et al. (2013) observed that maize near-isogenic lines resistant against Gibberella stalk rot also showed less severe seedling root browning and shrinkage after *F. graminearum* inoculation. This suggests that resistance against Gibberella stalk rot and *F. graminearum* seedling blight, both diseases in vegetative tissues, are likely controlled by similar genetic and physiological mechanisms, and are distinct from those affecting Gibberella ear rot resistance. This is consistent with the known influence of tissue type and developmental stage on the interactions between maize and *F. graminearum* (Zhang et al. 2016).

Genetic improvement of *F. graminearum* resistance depends on accurate quantitative assessment of resistant and susceptible phenotypes. Here we show that *F. graminearum* inoculation of maize seedlings not only inhibits root growth, but can also promote shoot elongation in a dosage- and genotype-dependent manner. Shoot elongation in response to *F. graminearum* root infection is significantly correlated with seedling survival, suggesting that these traits could be used to assess seedling resistance, as well as serving as a quantitative measure of the aggressiveness of different *F. graminearum* isolates.

Results

Fusarium graminearum inoculation at low concentration induces shoot elongation and root reduction in maize seedlings.

Different concentrations of spore suspensions have been used to study natural variation in maize seedling responses to *F. graminearum* inoculation (Ye et al., 2013; Kuhnem et al., 2015). To determine the optimal concentration of *F. graminearum* spores that leads to the largest change in plant morphology, two concentration gradient experiments were performed with Gz014, a field-collected *F. graminearum* strain (Table S2), and two commercial maize hybrids with differing seedling vigor levels (Hybrids IV and V in Table S1, as assessed by the provider, Blue River Organic Seed Company).

At 1.2 x 10^6 spores/mL, a concentration that is comparable to a previous study, the seedling stunting phenotype that had been reported was not observed in either of the tested maize hybrids (Kuhnem et al. 2015). Instead, the hybrid with high seedling vigor showed no significant

change in seedling height, whereas the moderately vigorous hybrid seedlings grew significantly taller when inoculated with *F. graminearum* spores. At lower concentrations, fungus-induced shoot elongation became obvious in both highly and moderately vigorous hybrid seedlings (Figure 2.1A). Underground, *F. graminearum* infection significantly reduced root growth in the moderately vigorous hybrid, with no evidence of a dosage response. This fungus-induced change was not observed in the highly vigorous hybrid, which showed no significant change in total root length at all tested doses (Figure 2.1B).

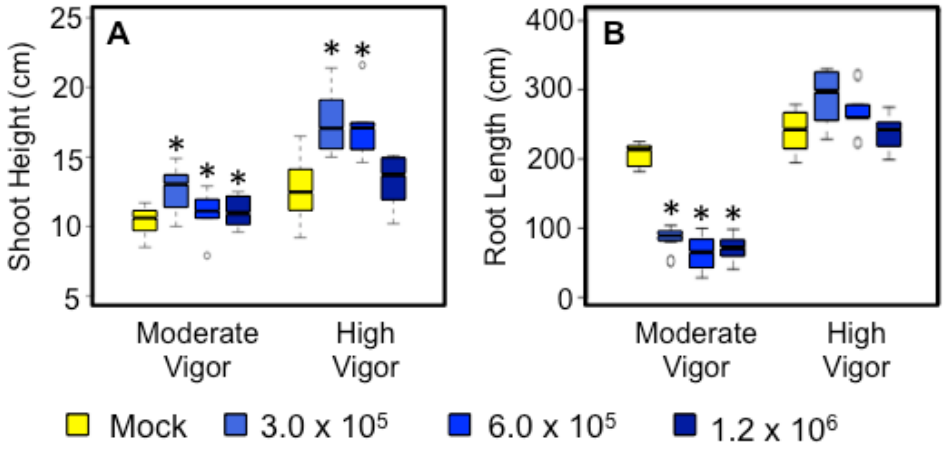


Figure 2.1. *Fusarium graminearum* inoculation induces shoot elongation and root reduction in maize seedlings in a dosage- and host genotype-dependent manner. Shoot height (A) and total root length (B) measurements on seedlings inoculated at each fungal spore concentration were compared to mock-treated seedlings with Dunnett's tests relative to mock-infected controls (N = 6 for each group; * $p < 0.05$). Group means are indicated by the black bars, and the interquartile range by the upper and lower edges of the boxes. Outliers extending more than 1.5x of the interquartile range are indicated by individual circles, and the whiskers show the range of each group excluding the outliers. Fungal strain Gz014 was used for this experiment. All measurements were taken 9 days post-inoculation. Fungal spore concentration is in counts/milliliter.

For both fungus-induced growth responses, inoculation with the lowest fungal spore concentration (3 x 10^5 count/mL) resulted in a response that was stronger than or comparable to that observed with the two higher inoculum concentrations. Therefore, the 3 x 10^5 spores/mL *F. graminearum* dose was used for all subsequent experiments.

***Fusarium graminearum*-induced shoot elongation and root reduction vary independently across maize inbred lines.**

To confirm the *F. graminearum*-induced shoot elongation and root reduction observed in the concentration gradient experiment, the artificial inoculation experiment was repeated with the reference maize genetic inbred line B73 with the same fungal isolate following the same

protocol. Consistently, the same fungus-induced growth responses were observed (Figure 2.2). To investigate how *F. graminearum*-induced seedling shoot elongation and root reduction may be related, seedlings of fifteen maize inbred lines were inoculated with Gz014 or mock inoculum and their shoot and root growth were quantified 9 days after inoculation. Although *F. graminearum* inoculation reduced total root length in most inbred lines by approximately 50%, no significant difference was observed between mock- and *F. graminearum*-inoculated NC350, NC358, and Mo17 seedlings, (Figure 2.3A).

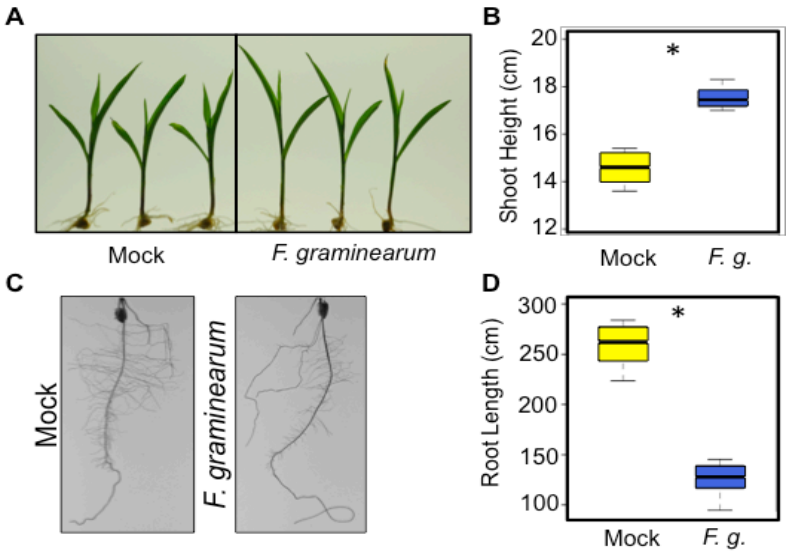


Figure 2.2. *Fusarium graminearum* inoculation induces shoot elongation and root reduction in maize inbred line B73. Representative photos of shoots (A) and roots (C) mock- and fungus-inoculated B73 seedlings were taken 9 days post-inoculation. Shoot height (B) and total root length (D) were measured on the same day and compared with Student's *t*-tests (N = 6 for each group; *p < 0.05). Group means are indicated by the black bars, and the interquartile range by the upper and lower edges of the boxes. Outliers extending more than 1.5x of the interquartile range were indicated by individual circles, and the whiskers showed the range of each group excluding the outliers. Fungal strain Gz014 was used for this experiment. Spore concentration = 3 x 10^5 counts/milliliter.

In the same experiment, *F. graminearum*-induced shoot elongation also varied, ranging from 0 to over 25% increase (Figure 2.3B). Noticeably, the two fungus-induced growth responses do not correlate with one another across the fifteen tested inbred lines tested (Pearson's $R^2 = 0.0079$, $p > 0.05$). Two inbred lines showing no significant *F. graminearum*-induced root reduction, NC350 and NC358, grew 16% and 24% taller, respectively, after *F. graminearum* inoculation. Conversely, Tx303 only showed obvious root reduction (40%) but no significant shoot elongation (Figure 2.3). Inbred line Mo17 showed neither of the tested fungus-induced growth responses.

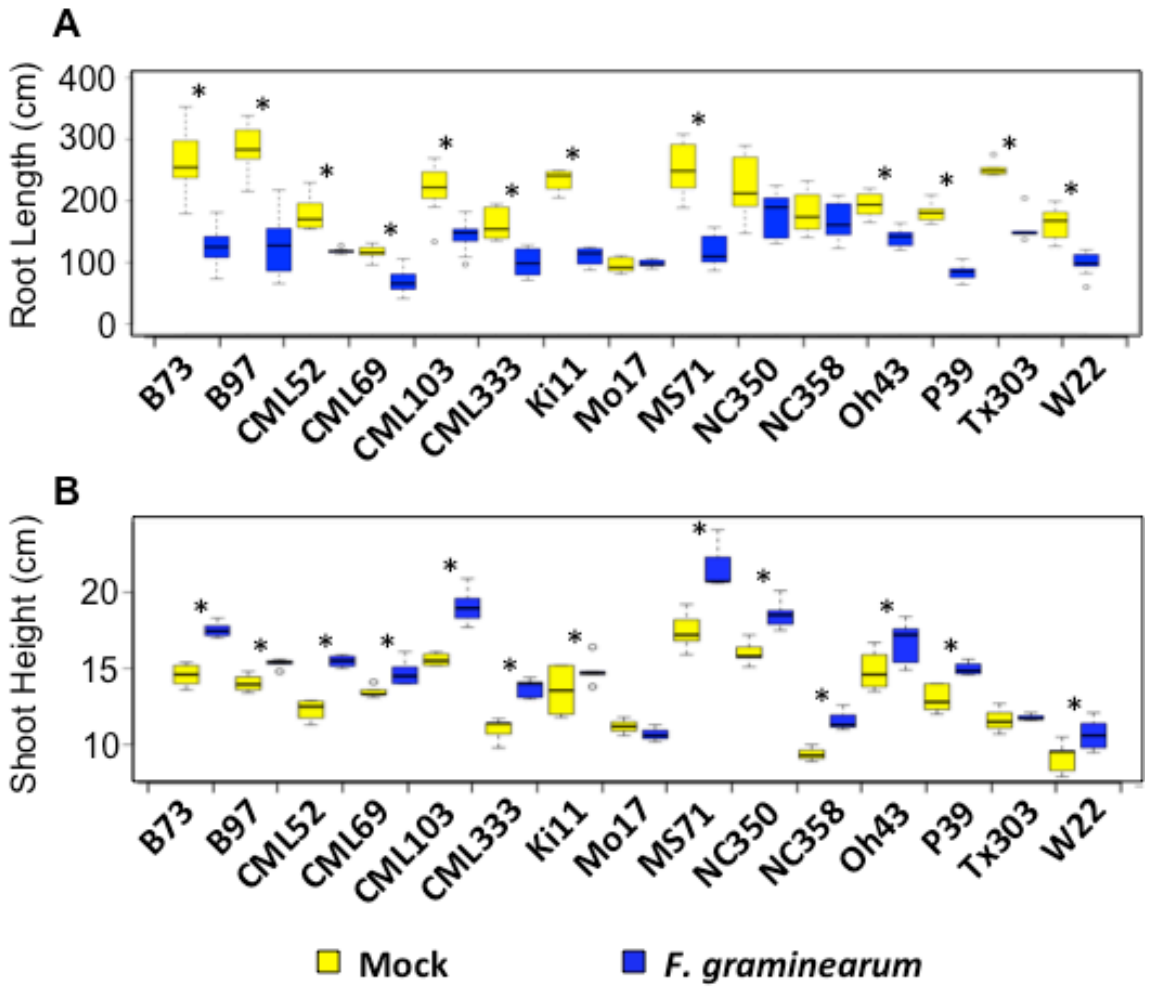


Figure 2.3. *Fusarium graminearum*-induced growth responses in maize seedlings vary across maize inbred lines. Root length (A) and shoot height (B) of each maize inbred line inoculated with *F. graminearum* spores (yellow) were compared to measurement from a corresponding mock treatment group (blue) with Student's *t*-test (* $p < 0.05$). Group means are indicated by the black bars, and the interquartile range by the upper and lower edges of the boxes. Outliers extending more than 1.5x of the interquartile range are indicated by individual circles, and the whiskers show the range of each group excluding the outliers. All measurements were taken at 9 days post-inoculation. Fungal strain Gz014 was used for this experiment. Spore concentration = 3 x 10^5 counts/milliliter.

***Fusarium graminearum*-induced growth responses and *F. graminearum* seedling blight severity are primarily determined by the fungal genotype.**

To further survey the natural variation in *F. graminearum*-induced seedling shoot elongation and root reduction, the artificial inoculation experiment was expanded to the nine commercial maize hybrids, including the two used in the concentration gradient experiment described above (Table S2.1), and three *F. graminearum* isolates (Table S2.2). Both *F. graminearum*-induced growth responses were highly variable across this panel (Figure 2.4). Two-way analysis of variance

(ANOVA) performed on both traits showed that fungal isolate, rather than maize maturation time or provider-assessed seedling vigor, is the main source of assay variance (Table S2.3).

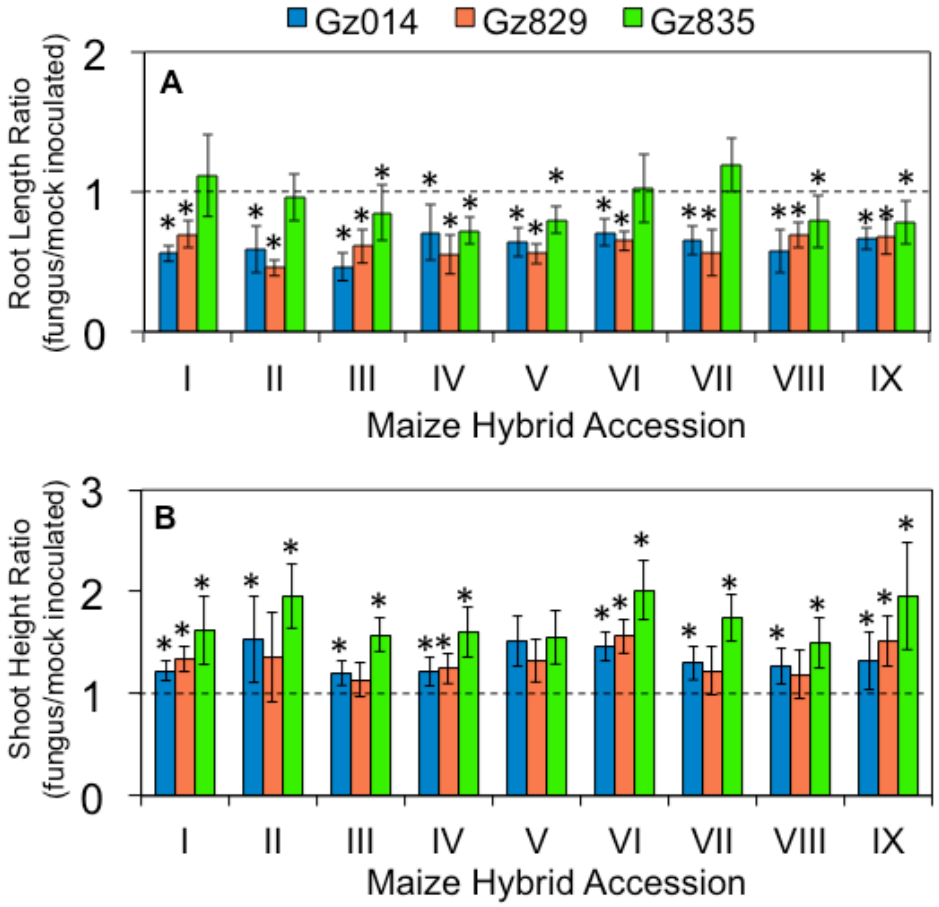


Figure 2.4. *Fusarium graminearum*-induced growth responses in maize seedlings vary with both plant and fungal genotypes. Nine maize hybrid lines were inoculated with three *F. graminearum* isolates (Gz014, Gz829, and Gz835). Total root length (A) and shoot height (B) of each maize hybrid-fungus isolate pair were compared to measurement from a corresponding mock treatment group with Student's *t*-test (* $p < 0.05$). All measurements were taken at 9 days post-inoculation. Fungal spore concentration = 3 x 10^5 count/milliliter.

After quantifying *F. graminearum*-induced growth responses, the fungus-infected seedlings were kept under the same growth conditions, and *F. graminearum* seedling blight severity was quantified as the percentage of seedlings surviving in each maize hybrid population fourteen days after inoculation. As in the case of the morphological data, ANOVA of the seedling survival rate data identified the *F. graminearum* isolate as the most important contributor. Whereas both Gz014 and Gz829 infection left less than 50% of inoculated seedlings surviving after fourteen days, more than 85% of Gz835-inoculated seedlings survived in the

same time period (Figure 2.5). The maturation time and seedling vigor of maize hybrids were not significant sources of variance (Table S3).

***Fusarium graminearum*-induced shoot elongation is positively correlated with *F. graminearum* seedling blight severity.**

To investigate the relationship between *F. graminearum*-induced growth responses and *F. graminearum* seedling blight severity, linear regression analysis was performed on the extent of these changes and seedling survival. This showed that *F. graminearum*-induced seedling elongation has a significant negative correlation with seedling survival rate at 14 days after inoculation (Figure 6A; Pearson's $R^2 = 0.244$; $p < 0.005$). Thus, stronger fungus-induced shoot elongation may be predictive of more severe *F. graminearum* seedling blight. Unexpectedly, the extents of *F. graminearum*-induced shoot elongation showed strong positive correlation with root length (Figure 6B; Pearson's $R^2 = 0.676$; $p < 0.005$), and a significant negative correlation between root reduction and seedling survival rate (Figure 6C; Pearson's $R^2 = 0.263$; $p < 0.005$). This leads to the counter-intuitive inference that resistance to fungus-induced root growth inhibition co-occurs with stronger shoot elongation and more severe *F. graminearum* seedling blight.

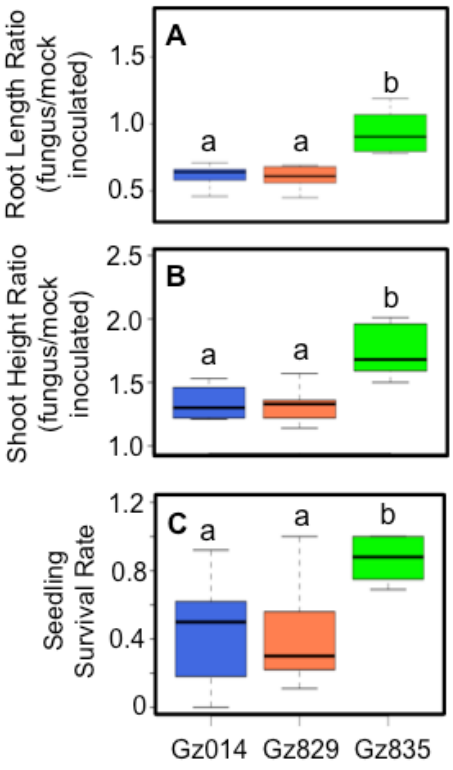


Figure 2.5. *Fusarium graminearum*-induced growth responses and seedling blight severity vary significantly between fungal isolates. The extent of *F. graminearum*-induced root reduction and shoot elongation at 9 days post-inoculation (A & B) and survival rates of maize seedlings inoculated with different *F. graminearum* isolates at 14 days post-inoculation (C) were compared using two-way analysis of variance with Tukey's HSD. Group means are indicated by the black bars, and the interquartile range by the upper and lower edges of the boxes. Outliers extending more than 1.5x of the interquartile range were indicated by individual circles, and the whiskers showed the range of each group excluding the outliers. Significantly different groups are denoted with different letters. Fungal spore concentration = 3×10^5 count/milliliter. Ratios refer to fungus/mock-inoculated plants.

Discussions

Under field conditions, maize seedlings infected by *F. graminearum* as well as other fungal and oomycete pathogens are often characterized by stunted growth and dehydration in the

aboveground tissues before death. Similar seedling stunting symptoms have also been observed in a previous artificial inoculation test with *F. graminearum* spores under laboratory conditions (Kuhnem et al., 2015). However, in our concentration gradient experiment with the same fungal isolates, low dosage *F. graminearum* inoculation on maize seedling root induced shoot elongation (Figure 1A; Figure 2A). Based on the natural variation in this *F. graminearum*-induced response across different maize inbred lines and hybrids reported in this study (Figure 3A; Figure 4A), it is likely that the inconsistent observations between studies could arise from inoculum concentration, maize genotype, or unintended differences between experimental conditions.

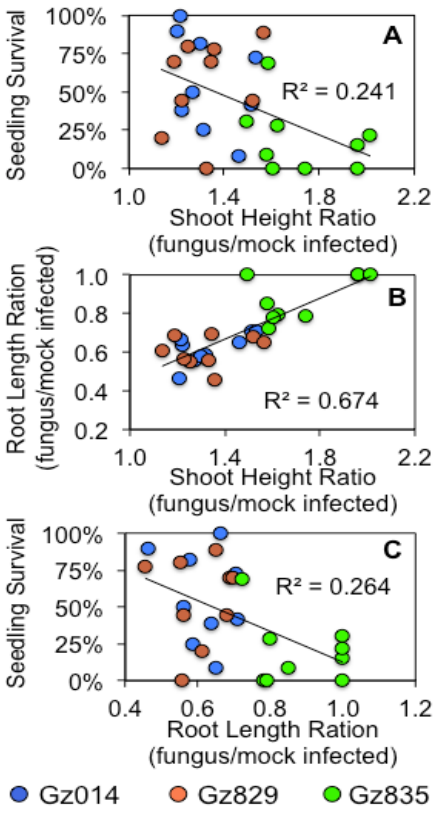


Figure 2.6. *Fusarium graminearum*-induced growth parameter changes in roots and shoots, and seedling blight severity, which are significantly correlated with each other. Each mark represents a maize hybrid-fungal isolate pair, with different fungal isolates represented by different colors. Root length and seedling height were measured at 9 days after infection. Survival was assessed at 14 days after infection

Interestingly, the observed *F. graminearum*-induced shoot elongation phenotype is reminiscent of the foolish seedling disease in rice, which is caused by a related fungal pathogen *Gibberella fujikuroi*. Rice seedlings affected by foolish seedling disease demonstrate abnormal seedling elongation resulting from the production of gibberellic acid (GA), a growth-promoting phytohormone, *G. fujikuroi* (Yabuta and Sumiki 1938). Fungus-derived GA has been hypothesized to be a virulence factor by negatively influencing the signaling transduction pathway of the immunity-regulating phytohormone, jasmonic acid (Navarro et al. 2008). However, unlike *G. fujikuroi, F. graminearum* is not known to produce GA, and its genome does not contain biosynthetic genes required for GA biosynthesis (Cuomo et al. 2007). However, we cannot rule out the possibility that *F. graminearum* infection induces endogenous maize GA production.

Consistent with previous reports (Ye et al. 2013; Kuhnem et al., 2015), *F. graminearum* infection induced maize seedling root growth reduction in a host genotype-dependent manner

(Figure 1B; Figure 2B; Figure 3B; Figure 4B). Application of a computer image-processing tool, RootReader 2D (Clark et al. 2013), in our study enabled more accurate, quantitative, and efficient measurement of fungus-induced growth response, and its natural variation with host and pathogen genotype. Fungus-induced root reduction and shoot elongation measurements from the same experiment on fifteen maize inbred lines showed no correlation in the extent of these two traits, with some inbred lines showing only one or the other fungus-induced change (Figure 3A&B). This observation suggests that these two responses may be governed by distinct physiological mechanisms.

As one of the most widespread fungal phytopathogens, *F. graminearum* is known to be highly diverse, and different strains are highly variable in their aggressiveness (Backhouse 2014; Boutigny et al. 2011; Lee et al. 2010; Sampietro et al. 2011; Wang et al. 2011). Our survey of nine commercial maize hybrids and three fungal isolates (Figure 4 & 5) is consistent with prior observations that the prevalence of *F. graminearum*-related diseases is highly variable across multiple locations, depending on the aggressiveness of the predominant isolate and environmental permissiveness (Backhouse 2014; Boutigny et al. 2011; Lee et al. 2010; Sampietro et al. 2011; Wang et al. 2011).

Unexpectedly, maize hybrid-fungal isolate pairs with stronger fungus-induced root growth inhibition tended to show a lower level of induced shoot elongation (Figure 6C) and better survival (Figure 6B). This was surprising, because severe reduction in root growth was expected to significantly impair the seedlings' ability to acquire water and nutrients from the soil, which could lead to seedling death. Two-way ANOVA and a closer look at the data distribution suggest that these correlative relationships are primarily driven by differences between the fungus isolates. Specifically, the isolate that induced significantly stronger shoot elongation, Gz835, induced less root reduction and seedling death (Figure 5). This suggests that these fungus-induced growth responses can be used for quantitative assessment of the aggressiveness of *F. graminearum* isolates, and that different *F. graminearum* isolates may have distinct pathogenic mechanisms when interacting with maize seedlings. Further research with maize seedlings growing in agricultural fields will be required to determine whether *F. graminearum*-induced shoot elongation can serve as a reliable predictor of fungal sensitivity in commercially grown maize.

Materials and Methods

Plant and fungal materials. Fifteen parental lines from the maize nested association mapping population (McMullen et al. 2009) were chosen for *F. graminearum* infection assays. For a broader survey, nine organic commercial maize hybrids were kindly provided by the Blue River Organic Seed Company (Table S2.1). The maize hybrids were selected to cover different maturation time and provider-assessed seedling vigor levels. The accession numbers of the tested hybrids are masked per the provider's request. Three field-collected *F. graminearum* isolates (Table S2.2) that have shown different levels of aggressiveness in a previous laboratory test (Kuhnem et al. 2015) were used for inoculations (Table S2.2). Fungal chemotype, namely the type of major mycotoxin accumulated, was not considered during isolate selection, since it was not significantly associated with aggressiveness in the previous test.

Plant growth and fungal inoculation methods. Maize seeds were germinated in moisturized rolls of germination paper after surface sterilization with 10% household bleach for 30 minutes. Germinated seedlings were transplanted to 7.5 cm × 7.5 cm plastic pots with Turface ® MVP ® calcined clay (Profile Products LLC, Buffalo Grove, IL) when their primary roots reached approximately 9 cm, which would take 3-5 days depending on the specific maize genotype. For inoculation experiments, seedling roots were immersed in *F. graminearum* spore suspension or mock solution (0.03% Phytagar) for one hour prior to transplanting. Fungal spore suspensions were freshly prepared by flooding and scraping 7-day-old fungal cultures maintained on Potato Dextrose Agar plates under black lights with 0.03% Phytagar suspension. The spore concentration was estimated with a hemocytometer (Hausser Scientific, Horsham, PA) and light microscopy. Fungal hyphae fragments were observed in spore suspension by light microscopy.

For the concentration gradient experiment, twenty-four seedlings of each maize hybrid were grown, and six were inoculated with each tested fungal spore concentration and mock inoculum. For each pair of maize hybrid and fungal isolate included in the diversity screening, at least 15 seedlings were inoculated with fungal spores, and 5 were mock-inoculated. Transplanted seedlings were kept at 26°C (Day)/23°C (Night), 60% humidity, and a 16 hour long-day light cycle. Seedlings inoculated with different *F. graminearum* isolates or mock inoculum were kept in separated growth chambers set to identical growth conditions.

Phenotypic data collection and analysis. The heights of fungus- and mock-inoculated seedlings were measured nine days post-inoculation. On the same day, seedlings were removed from pots

to measure their total root length using an image analysis algorithm implemented in the RootReader 2D software (Clark et al. 2013). After measurement, seedlings were re-potted in the same Turface particles. At fourteen days post-inoculation, the number of surviving seedlings within each fungus-inoculated maize hybrid population was counted as a measurement of *F. graminearum* seedling blight severity. Seedlings were considered dead when they broke at the soil line, with browning and rot at the breakpoint.

For seedling height and total root length, an average ratio of mock- and fungus-inoculated seedlings for each maize hybrid-fungus isolate combination was calculated to shown variation in these fungus-induced responses. Measurements from fungus- and mock-inoculated seedlings were compared using Student's *t*-tests. Average ratios were rounded to one when measurements from the two groups were not significantly different from one another. For two-way ANOVA, maturation times of maize hybrids were grouped into four bins, as indicated in parentheses in Table S2.1. The average ratios of seedling height and total root length between fungus- and mock-treated seedlings were used as quantitative measurements to calculate linear correlation with seedling survival rate measured at 14 days after infection.

References

Ali ML, Taylor JH, Jie L, Sun G, William M, Kasha KJ, Reid LM, Pauls KP (2005) Molecular mapping of QTLs for resistance to Gibberella ear rot, in corn, caused by *Fusarium graminearum*. Genome 48 (3):521-533. doi:10.1139/g05-014

Arunachalam C, Doohan FM (2013) Trichothecene toxicity in eukaryotes: Cellular and molecular mechanisms in plants and animals. Toxicol Lett 217 (2):149-158. doi:10.1016/j.toxlet.2012.12.003

Backhouse D (2014) Global distribution of *Fusarium graminearum*, *F. asiaticum* and *F. boothii* from wheat in relation to climate. Eur J Plant Pathol 139 (1):161-173. doi:10.1007/s10658-013-0374-5

Boutigny AL, Ward TJ, Van Coller GJ, Flett B, Lamprecht SC, O'Donnell K, Viljoen A (2011) Analysis of the *Fusarium graminearum* species complex from wheat, barley and maize in South Africa provides evidence of species-specific differences in host preference. Fungal Genet Biol 48 (9):914-920. doi:10.1016/j.fgb.2011.05.005

Brauner PC, Melchinger AE, Schrag TA, Utz HF, Schipprack W, Kessel B, Ouzunova M, Miedaner T (2017) Low validation rate of quantitative trait loci for Gibberella ear rot resistance in European maize. Theor Appl Genet 130 (1):175-186. doi:10.1007/s00122-016-2802-3

Chen Q, Song J, Du WP, Xu LY, Jiang Y, Zhang J, Xiang XL, Yu GR (2017) Identification, mapping, and molecular marker development for Rgsr8.1: a new quantitative trait locus conferring resistance to Gibberella stalk rot in maize (Zea mays L.). Front Plant Sci 8:1355. doi:10.3389/fpls.2017.01355

Clark RT, Famoso AN, Zhao KY, Shaff JE, Craft EJ, Bustamante CD, Mccouch SR, Aneshansley DJ, Kochian LV (2013) High-throughput two-dimensional root system phenotyping platform facilitates genetic analysis of root growth and development. Plant Cell Environ 36 (2):454-466. doi:10.1111/j.1365-3040.2012.02587.x

Cuomo CA, Guldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma LJ, Baker SE, Rep M, Adam G, Antoniw J, Baldwin T, Calvo S, Chang YL, Decaprio D, Gale LR, Gnerre S, Goswami RS, Hammond-Kosack K, Harris LJ, Hilburn K, Kennell JC, Kroken S, Magnuson JK, Mannhaupt G, Mauceli E, Mewes HW, Mitterbauer R, Muehlbauer G, Munsterkotter M, Nelson D, O'Donnell K, Ouellet T, Qi W, Quesneville H, Roncero MI, Seong KY, Tetko IV, Urban M, Waalwijk C, Ward TJ, Yao J, Birren BW, Kistler HC (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science 317 (5843):1400-1402. doi:10.1126/science.1143708

Kebede AZ, Woldemariam T, Reid LM, Harris LJ (2016) Quantitative trait loci mapping for Gibberella ear rot resistance and associated agronomic traits using genotyping-by-sequencing in maize. Theor Appl Genet 129 (1):17-29. doi:10.1007/s00122-015-2600-3

Kuhnem PR, Del Ponte EM, Dong YH, Bergstrom GC (2015) *Fusarium graminearum* ssolates from wheat and maize in New York show similar range of aggressiveness and toxigenicity in cross-species pathogenicity tests. Phytopathology 105 (4):441-448. doi:10.1094/Phyto-07-14-0208-R

Lee SH, Lee J, Nam YJ, Lee S, Ryu JG, Lee T (2010) Population structure of *Fusarium graminearum* from maize and rice in 2009 in Korea. Plant Pathology J 26 (4):321-327. doi:10.5423/Ppj.2010.26.4.321

Ma C, Ma X, Yao L, Liu Y, Du F, Yang X, Xu M (2017) qRfg3, a novel quantitative resistance locus against Gibberella stalk rot in maize. Theor Appl Genet 130 (8):1723-1734. doi:10.1007/s00122-017-2921-5

Maresca M (2013) From the gut to the brain: journey and pathophysiological effects of the food-associated trichothecene mycotoxin deoxynivalenol. Toxins 5 (4):784-820. doi:10.3390/toxins5040784

McMullen M, Bergstrom G, De Wolf E, Dill-Macky R, Hershman D, Shaner G, Van Sanford D (2012) A unified effort to fight an enemy of wheat and barley: Fusarium head blight. Plant Dis 96 (12):1712-1728. doi:10.1094/Pdis-03-12-0291-Fe

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, Brown P, Browne C, Eller M, Guill K, Harjes C, Kroon D, Lepak N, Mitchell SE, Peterson B, Pressoir G, Romero S, Oropeza Rosas M, Salvo S, Yates H, Hanson M, Jones E, Smith S, Glaubitz JC, Goodman M, Ware D, Holland JB, Buckler ES (2009) Genetic properties of the maize nested association mapping population. Science 325 (5941):737-740. doi:10.1126/science.1174320

Mueller D (2016a) Corn disease loss estimates from the United States and Ontario, Canada - 2013. Purdue University Extension, Retrieved from http://cropprotectionnetwork.org/crop-loss-estimates/corn-disease-loss-estimates-2013/.

Mueller D (2016b) Corn disease loss estimates from the United States and Ontario, Canada - 2014. Purdue University Extension, Retrieved from http://cropprotectionnetwork.org/crop-loss-estimates/corn-disease-loss-estimates-2014/.

Mueller D (2016c) Corn disease loss estimates from the United States and Ontario, Canada - 2015. Purdue University Extension, Retrieved from http://cropprotectionnetwork.org/crop-loss-estimates/corn-disease-loss-estimates-2015/.
Mueller D (2017) Corn disease loss estimates from the United States and Ontario, Canada - 2016. Purdue University Extension, Retrieved from http://cropprotectionnetwork.org/crop-loss-estimates/corn-disease-loss-estimates-2016/.
Munkvold GP, White DG (2016) Compendium of Corn Diseases. Fourth edn. The American Phytopathological Society,
Navarro L, Bari R, Achard P, Lison P, Nemri A, Harberd NP, Jones JDG (2008) DELLAs control plant immune responses by modulating the balance and salicylic acid signaling. Curr Biol 18 (9):650-655. doi:10.1016/j.cub.2008.03.060
Sampietro DA, Diaz CG, Gonzalez V, Vattuone MA, Ploper LD, Catalan CAN, Ward TJ (2011) Species diversity and toxigenic potential of *Fusarium graminearum* complex isolates from maize fields in northwest Argentina. Int J Food Microbiol 145 (1):359-364. doi:10.1016/j.ijfoodmicro.2010.12.021
Trail F (2009) For Blighted Waves of Grain: *Fusarium graminearum* in the Postgenomics Era. Plant Physiol 149 (1):103-110. doi:10.1104/pp.108.129684
Wang JH, Ndoye M, Zhang JB, Li HP, Liao YC (2011) Population structure and Genetic diversity of the *Fusarium graminearum* species complex. Toxins 3 (8):1020-1037. doi:10.3390/toxins3081020
Yabuta T, Sumiki Y (1938) On the crystal of gibberellin, a substance to promote plant growth. Journal of the Agricultural Chemical Society of Japan 14 (1526)
Yang Q, Yin GM, Guo YL, Zhang DF, Chen SJ, Xu ML (2010) A major QTL for resistance to Gibberella stalk rot in maize. Theor Appl Genet 121 (4):673-687. doi:10.1007/s00122-010-1339-0
Ye JR, Guo YL, Zhang DF, Zhang N, Wang C, Xu ML (2013) Cytological and molecular characterization of quantitative trait locus qRfg1, which confers resistance to Gibberella stalk rot in maize. Mol Plant Microbe In 26 (12):1417-1428. doi:10.1094/Mpmi-06-13-0161-R
Zhang DF, Liu YJ, Guo YL, Yang Q, Ye JR, Chen SJ, Xu ML (2012) Fine-mapping of qRfg2, a QTL for resistance to Gibberella stalk rot in maize. Theor Appl Genet 124 (3):585-596. doi:10.1007/s00122-011-1731-4
Zhang Y, He J, Jia LJ, Yuan TL, Zhang D, Guo Y, Wang Y, Tang WH (2016) Cellular Tracking and gene profiling of *Fusarium graminearum* during maize stalk rot disease development elucidates its strategies in confronting phosphorus limitation in the host apoplast. PLoS Pathog 12 (3):e1005485. doi:10.1371/journal.ppat.1005485

CHAPTER THREE

ETHYLENE SIGNALING REGULATES NATURAL VARIATION IN THE ABUNDANCE OF ANTIFUNGAL ACETYLATED DIFERULOYLSUCROSES AND *FUSARIUM GRAMINEARUM* RESISTANCE IN MAIZE SEEDLING ROOTS[1, 2]

Abstract

The production of defensive specialized metabolites is required for pathogen resistance in maize (*Zea mays*) and other plants. Through comparative metabolomic analysis of inbred lines B73 and Mo17, we identified two maize metabolites, smilaside A (3,6-diferuloyl-3′,6′-diacetylsucrose) and smiglaside C (3,6-diferuloyl-2′,3′,6′-triacetylsucrose), that are induced by infection with *Fusarium graminearum* and other fungal pathogens. Smilaside A, which shows greater *in vitro* antifungal activity than smiglaside C*,* is more abundant in the *F. graminearum*-resistant Mo17 maize inbred line than in the sensitive B73 inbred line. Quantitative trait mapping with B73 × Mo17 recombinant inbred lines and near-isogenic lines, combined with whole transcriptome profiling, shows that elevated expression of the ethylene receptor gene *ETHYLENE INSENSITIVE 2* (*ZmEIN2*) increases the smilaside A/smiglaside C ratio. This regulatory function of ethylene is confirmed by changes in the abundance of smilaside A and smiglaside C in maize roots after exogenous addition of the ethylene precursor 1-aminocyclopropane-1-carboxylate (1-ACC), analysis of a 1-ACC synthase (*Zmacs2-1 Zmacs6*) mutant, and inhibition of ethylene synthesis with 1-methylcyclopropene (1-MCP). The ethylene-induced increase in smiglaside C abundance relative to the more strongly antifungal smilaside A, as well decreased *F. graminearum* growth after 1-MCP treatment, demonstrates that ethylene negatively regulates of *F. graminearum* resistance in maize.

[1] All work in this chapter is in revision for publication in *New Phytologist* as of July 18th, 2018

[2] Shaoqun Zhou designed and performed most of experiments. Ying Zhang and Dr. Frank Schroeder collected and analyzed NMR spectroscopy data of purified compounds. Dr. Karl Kremling and Dr. Ed Buckler collected transcriptomics data. Dr. Yezhang Ding and Dr. Eric Schmelz performed the fungal pathogen induction experiment not involving *Fusarium graminearum*. John Bennett and Dr. Michael Kolomiets performed ergosterol measurement experiment with the *Zmacs2-1 Zmacs6* double mutants. Justin Bae, Dean Kim, and Hayley Ackerman assisted in other experiments.

Introduction

Plants in natural and manmade ecosystems are continuously exposed to microbial pathogens. Specialized metabolic pathways that give rise to diverse arsenals of bioactive defense compounds allow plants to efficiently fend off pathogen attacks. The significance of plant specialized metabolism in agriculture is exemplified by the association of specific biosynthetic genes with resistance against insect pests and phytopathogens (Meihls *et al.*, 2013; Handrick *et al.*, 2016; Yang *et al.*, 2017). Such studies highlight the potential of enlisting naturally occurring specialized metabolites in crop species to enhance quantitative disease resistance.

In North America, maize (*Zea mays*) is the most important agricultural crop, with over 13 billion bushels produced per annum, of which approximately 10% is lost to disease (Mueller, 2016a; Mueller, 2016b; Mueller, 2017). Maize is also known for its great genetic diversity, involving both nucleotide polymorphisms and structural genomic variation (Buckler *et al.*, 2006; Jiao *et al.*, 2017). The genetic architecture of disease resistance in maize has been investigated extensively using publicly available genetic resources (Mideros *et al.*, 2012; Olukolu *et al.*, 2014; Benson *et al.*, 2015). Compared to foliar and ear diseases, maize seedling diseases remain a relatively understudied area, even though in some years they can account for more yield loss than any single disease in the aboveground tissues (Mueller, 2016a). This may be due to the fact that experimental methods developed for large scale screening of diseases in aboveground tissues, such as controlled pathogen inoculation and visual symptom scoring, are difficult to apply to seedling diseases under field conditions.

Fusarium graminearum is one of the most common causal pathogens of maize seedling disease in the northern temperate zone. In the field, it overwinters on crop residue as thickened hyphae and produces asexual conidia that infect germinating seedlings roots or mesocotyls. Depending on the developmental stage and infection site, *F. graminearum* can also cause root rot, stem rot, and ear rot in maize (Munkvold & White, 2016). Previous research on maize-*F. graminearum* interactions has primarily focused on ear rot, with results consistently suggesting that resistance against this disease is most likely controlled by numerous small-effect quantitative trait loci (QTL) that are influenced by experimental methods and genetic backgrounds (Ali *et al.*, 2005; Kebede *et al.*, 2016; Brauner *et al.*, 2017). Furthermore, transcriptomic studies in maize and wheat show that host-*F. graminearum* interactions are significantly influenced by host tissue

types, suggesting that the QTL associated with *F. graminearum* ear rot resistance probably will not confer resistance in seedling roots (Kazan *et al.*, 2012; Zhang *et al.*, 2016).

Compared to QTL identified from *F. graminearum* ear rot studies, factors contributing to *F. graminearum* stalk rot resistance may be more relevant to infections of seedling roots. For instance, near-isogenic lines (NILs) selected based on stalk rot resistance phenotypes also showed significant differences in primary root symptoms after controlled inoculation (Ye *et al.*, 2013). *Fusarium graminearum* infection in maize stalks induces production of specialized metabolites with antifungal activities (Huffaker *et al.*, 2011; Schmelz *et al.*, 2011). Additionally, comparative and correlative studies have identified constitutive phytoanticipins that were associated with *F. graminearum* resistance. For example, an *F. graminearum*-resistant NIL was found to accumulate significantly higher phenolic acids in its seedling roots compared to susceptible relatives. Interestingly, these differences disappear after *F. graminearum* infection, primarily due to fungus-induced reduction of defenses in the resistant NILs (Ye *et al.*, 2013). The same compounds also have been identified as metabolites related to *F. graminearum* resistance in other crop species, and were shown to inhibit fungal growth *in vitro* (Bollina *et al.*, 2010; Ponts *et al.*, 2011). Taken together, these studies indicate that specialized metabolites in maize seedling roots play a significant role in resistance against *F. graminearum.*

Here we describe a comparative metabolomics approach using maize inbred lines B73 and Mo17 to identify previously unknown maize antifungal compounds. These experiments led to the identification of two acetylated diferuloylsucroses, one of which demonstrated significant fungal growth inhibition *in vitro* at a physiologically relevant concentration. Genetic mapping, analysis of mutants, and physiological experiments demonstrated that accumulation of acetylated diferuloylsucroses is promoted by ethylene production and fine-tuned by ethylene sensitivity in maize.

Results

Fusarium graminearum infection leads to root growth inhibition and metabolic reconfiguration

Inoculation of maize genotype B73 with *F. graminearum* under controlled growth conditions, significantly reduced seedling root growth (Figure S3.1a,b). Using this assay, we screened the 26 parental lines of the maize nested association mapping population (McMullen *et al.*, 2009), as

well as inbred lines Mo17 and W22. This showed that B73 was among the most susceptible inbred lines, with root growth being reduced by over 50%. In contrast, Mo17 emerged as a

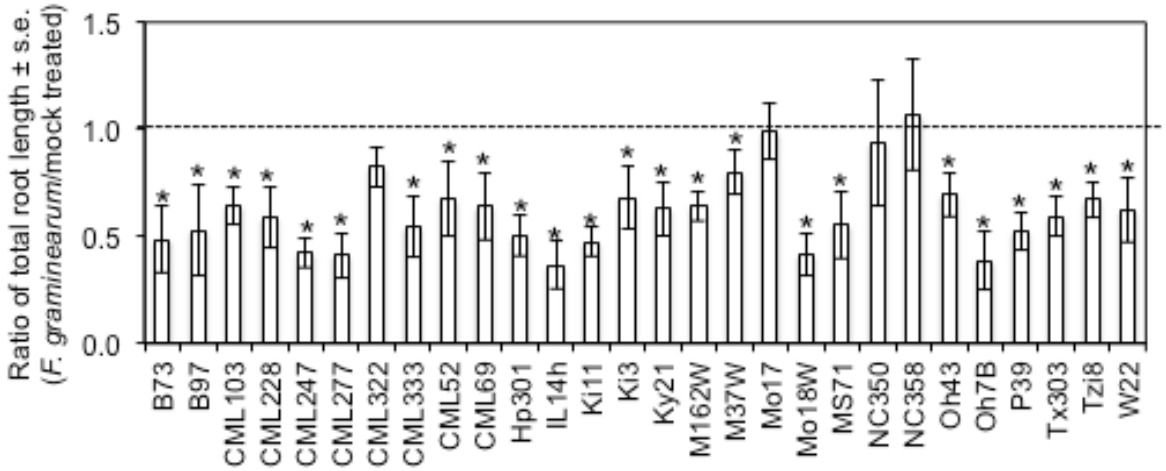


Figure 3.1. Natural variation in root growth inhibition by *F. graminearum*. The ratio of total root length of *F. graminearum*- and mock-inoculated maize seedlings, measured six days after inoculation, is shown. The dotted line denotes the expected ratio when there is no significant effect of *F. graminearum* infection. At least five mock- and fungus-inoculated individuals are measured for each genotype. *P < 0.05, Student's *t*-test for significant difference from 1.

potentially *F. graminearum*-resistant line, showing no significant change in root growth after inoculation (Figure 3.1 and Figure S3.1c). Further research was focused on the B73 *vs.* Mo17 comparison, due to the availability of genetic resources that included recombinant inbred lines (RILs) and near-isogenic lines (NILs) (Lee *et al.*, 2002; Eichten *et al.*, 2011). The higher fungal resistance of Mo17 was confirmed by lower expression of *FgTUB,* a *F. graminearum*-specific tubulin gene, lower levels of deoxynivalenol, a mycotoxin produced by *F. graminearum*, and fewer visible necrotic symptoms in the roots (Figure 3.2a-c).

We hypothesized that the contrasting *F. graminearum* resistance in B73 and Mo17 seedling roots could be attributed to differences in their constitutive and/or inducible biochemical defenses. Therefore, we performed a non-targeted comparative metabolomic analysis of B73 and Mo17 seedling roots, with and without *F. graminearum* inoculation. Consistent with the difference in *F. graminearum*-induced root growth reduction between these two inbred lines, we observed that more than three hundred mass features were significantly altered by *F. graminearum* infection of B73, but only twenty were altered in Mo17 in this experiment (Figure 3.2d).

Acetylated diferuloylsucroses contribute to *F. graminearum* resistance

To identify specific metabolites that could contribute to the contrasting *F. graminearum* resistance levels in B73 and Mo17, a separate non-targeted metabolomic experiment was performed to compare the constitutive metabolomes of B73 and Mo17 seedling roots, as well as

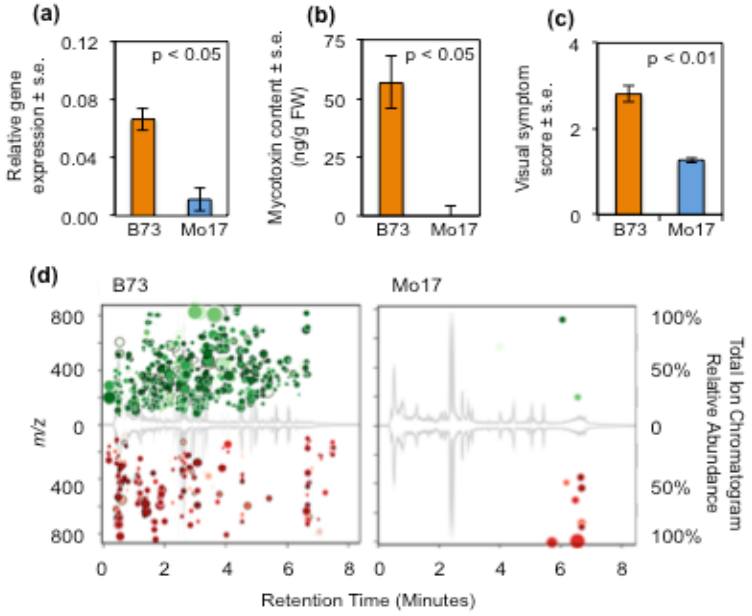


Figure 3.2. Mo17 is more resistant to *F. graminearum* than B73. Compared to B73, Mo17 seedling roots inoculated with *F. graminearum* demonstrate (a) lower expression of an *F. graminearum*-specific α-tubulin gene, mean +/- s.e. of N = 5, (b) lower content of deoxynivalenol at six-days-post-inoculation, mean +/- s.e. of N = 5, and (c) reduced symptoms on a 0-4 scale, with 0 = no symptoms and 4 = seedling death, mean +/- s.e. of N = 10. *P < 0.05 two-tailed unpaired Student's *t*-tests for A and B, and paired *t*-test for C. (d) Non-targeted metabolomics of *F. graminearum*-inoculated B73 and Mo17 seedling roots. For either genetic background, each mass feature that was significantly different between treatments (control vs. infected, N = 5; p < 0.05; fold change > 1.5) was plotted as a bubble on the total ion current chromatogram, with its size proportional to the fold change, and darker color representing a more significant change. Mass features induced by *F. graminearum* are shown in the upper half of each plot and *F. graminearum*-suppressed mass features are shown in the lower half.

mock- and *F. graminearum*-inoculated B73 seedling roots. This identified forty mass features that were both significantly affected by *F. graminearum* in the susceptible B73 seedling roots and constitutively different between B73 and Mo17 seedling roots (Table S3.1-3.3). Among these forty mass features, several represented specialized metabolites with known antifungal activity, including benzoxazinoids and phenylpropanoids (Bollina *et al.*, 2010; Ponts *et al.*, 2011; Kazan *et al.*, 2012).

Two mass features with mass-to-charge ratio (*m/z*) of 819.2321 and 777.2221 under negative electron spray ionization (ESI) mode, eluting at 6.11 and 5.61 minutes, respectively,

were significantly more abundant after *F. graminearum* infection of both B73 and Mo17 seedling roots. In all samples, the *m/z* 819 metabolite was much more abundant than the *m/z* 777 metabolite (Figure 3.3a,b). B73 contains significantly more of the *m/z* 819 metabolite than Mo17, both constitutively and after *F. graminearum* induction (Figure 3.3a). In contrast, the *m/z* 777 metabolite was more abundant in Mo17 under both conditions (Figure 3.3b).

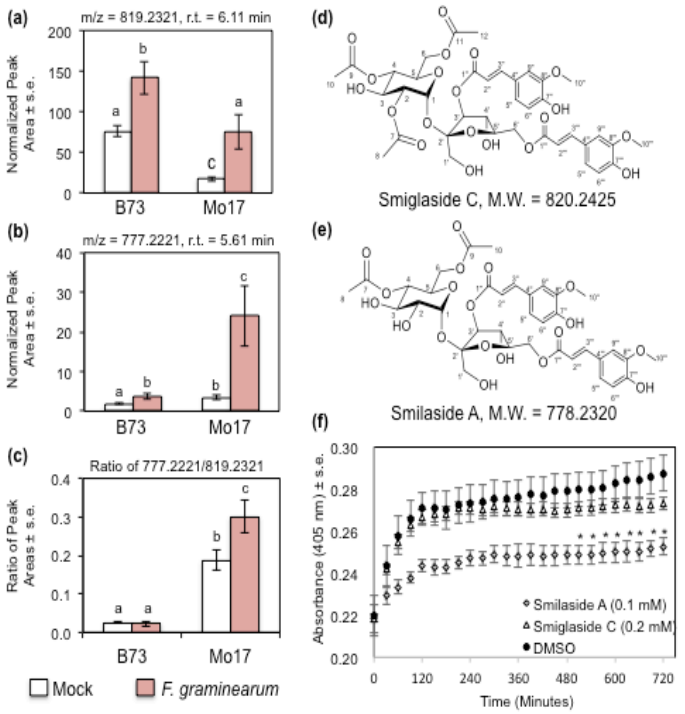


Figure 3.3. Antifungal metabolites smiglaside C and smilaside A are differentially induced by *F. graminearum* in B73 and Mo17. Abundance of metabolites with (a) *m/z* 819 and (b) *m/z* 777 was measured in negative electron spray ionization mode. (c) Ratio of the peak areas of the two metabolites. Mean +/- s.e. of N = 4. Different letters indicate significant differences, $P < 0.05$, ANOVA followed by Tukey's HSD test. Structures of (d) smiglaside C and (e) smilaside A were determined by LC-MS/MS and NMR. (f) Growth of an *F. graminearum* spore/hyphae suspension incubated with smilaside A, smiglaside C, or a dimethylsulfoxide (DMSO) solvent-only control. Fungal growth was monitored by absorbance at 405 nm. Mean +/- s.e. of N = 4, *$P < 0.05$, Student's *t*-test relative to the DMSO control at the same time point.

The *m/z* 819 metabolite was identified as 3,6-diferuloyl-2′,3′,6′-triacetylsucrose (Figure 3.3d), based on its phenylpropanoid-like UV absorbance profiles (Figure S3.2), tandem mass spectrometry (MS/MS; Figure S3.2), and nuclear magnetic resonance (NMR) spectroscopy (HSQC, HMBC, and dqfCOSY spectra; Table S3.4). Based on the MS/MS data and difference in exact mass, the *m/z* 777 metabolite was predicted to have one fewer acetyl group. This was confirmed by one dimensional proton NMR, which showed that the C2′ acetyl group was absent, and the compound was 3,6 diferuloyl-3′,6′-diacetylsucrose (Figure 3.3e). These two metabolites were previously identified as smilaside A (3,6 diferuloyl-3′,6′-diacetylsucrose) in *Smilax china* (Kuo *et al.*, 2005; Cho *et al.*, 2015) and smiglaside C (3,6-diferuloyl-2′,3′,6′-triacetylsucrose) in *Smilax glabra* (Chen *et al.*, 2000).

In addition to smilaside A and smiglaside C, other maize compounds co-elute with a UV-absorbance peak at 328 nanometers, characteristic of a phenylpropanoid moiety. These include

likely structural isomers of smilaside A and smiglaside C, with identical *m/z* ratio and different retention times, as well as possible monoacetylated (*m/z* 735.21) and tetraacetylated (*m/z* 861.24) diferuloylsucroses (Figure S3.3). However, the structures of these other maize metabolites were not confirmed, and their functions have not been investigated in this study. Notably, non-acetylated diferuloylsucrose (expected *m/z* ratio = 693.21 in negative ESI mode) was not detected.

The structural resemblance of smilaside A and smiglaside C suggested that they could be the substrate and product of an acetylation reaction, respectively. This acetylation is actively regulated in response to fungal infection, with *F. graminearum* infection inducing a significant increase in the smilaside A/smiglaside C ratio only in the resistant Mo17 seedlings, but not in the susceptible B73 ones (Figure 3.3c). This induced response suggested that smiglaside C and/or smilaside A play a role in maize biochemical defense against *F. graminearum*. *In vitro* fungal growth inhibition assays were conducted in liquid suspension culture using smiglaside C and smilaside A concentrations similar to those found in maize seedlings (Figure S3.4). Although it was tested at a lower concentration, smilaside A showed a more significant inhibition of *F. graminearum* growth *in vitro* than smiglaside C (Figure 3.3f). This was consistent with our earlier observation that the *F. graminearum*-resistant Mo17 seedlings have a higher constitutive smilaside A content, and further accumulated this compound upon fungus attack compared to the susceptible B73 seedlings (Figure 3.3a-c).

Genetic mapping of smiglaside C and smilaside A abundance identifies *ETHYLENE INSENSITIVE 2* as a candidate regulator

The constitutive difference in smilaside A and smiglaside C abundance between B73 and Mo17 seedling roots allowed us to investigate the genetic control of this natural variation. Composite interval mapping with seedling roots of 83 recombinant inbred lines (RILs) from the intermated B73 x Mo17 (IBM) population (Lee *et al.*, 2002; Wang, 2012), together with the two parental lines, showed that the most significant QTL for both metabolites is located at the same position on chromosome 3 (Figure 3.4a). In version 2 of the B73 reference genome (Refgen v2), this locus covers approximately 2.3 million base pairs, containing 90 annotated gene models (Schnable *et al.*, 2009; Anders *et al.*, 2015). Interestingly, the two QTL have opposite effects, with the B73 allele promoting constitutive smiglaside C abundance and reducing smilaside A abundance (Figure 3.4b,c). Since smilaside A and smiglaside C are likely the substrate-product

pair of an acetylation reaction, we hypothesize the mapped QTL regulates the efficiency of this reaction. We identified the same locus when mapping the smilaside A/smiglaside C ratio as a quantitative trait, with the B73 allele reducing the smilaside A/smiglaside C ratio (Figure S3.5).

To further confirm the role of the QTL in regulating the relative abundance of smiglaside C and smilaside A, we quantified the metabolites in B73-Mo17 near-isogenic lines (NILs) with reciprocal introgressions at this locus (Eichten *et al.*, 2011). Consistent with the RIL results, the smilaside A/smiglaside C ratio showed clear co-segregation with the genetic markers at the chromosome 3 QTL, with the NILs carrying the B73 allele having a lower ratio, irrespective of their genetic background (Figure 3.4d). A significant difference between NILs carrying either

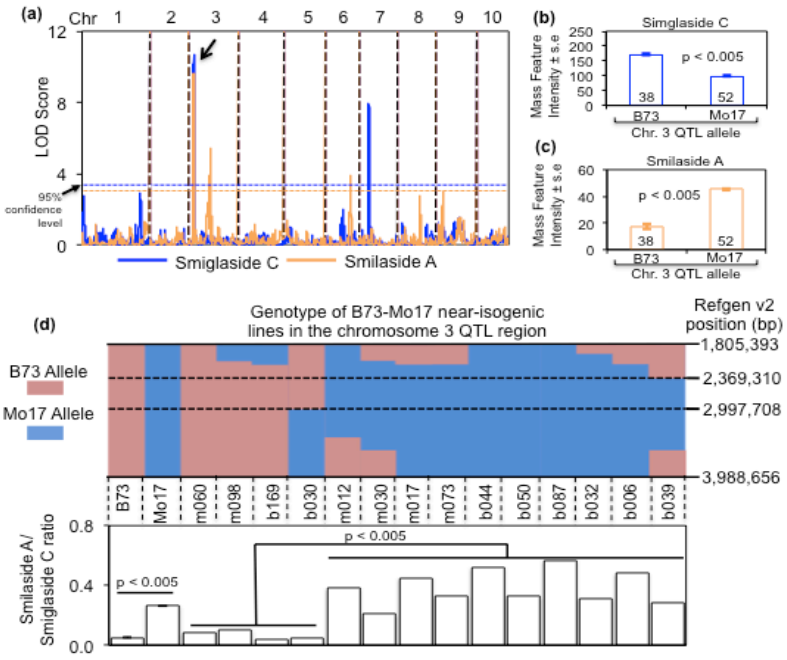


Figure 3.4. Smiglaside C and smilaside A share a QTL on chromosome 3 with opposite effects. (a) Composite interval mapping of smiglaside C (blue) and smilaside A (orange) in the seedling roots of B73 × Mo17 recombinant inbred lines (RILs) identified a significant QTL for both traits on chromosome 3 (indicated by an arrow). RILs with the B73 QTL allele on chromosome 3 have more smiglaside C (b) and less smilaside A (c) than those with the Mo17 allele. Numbers in bars are sample sizes. P values were determined with two-tailed Student's *t*-tests. (d) The smilaside A/smiglaside C ratio was calculated for B73, Mo17, and near-isogenic lines. The genetic background of the NILs is indicated by the initial letter of the line name, *e.g.* m060 has a Mo17 genetic background and b169 has a B73 genetic background. The smilaside A/smiglaside C ration is higher in Mo17 than in B73, mean +/- s.e. of N = 3, Student's *t*-test. NILs carrying the Mo17 allele at the Chromosome 3 QTL have a higher smilaside A/smiglaside C ratio than those with the B73 allele (Students *t*-test), irrespective of the overall genetic background.

allele was also observed for smilaside A but not smiglaside C (p = 0.075; Figure S6). Furthermore, due to additional recombination breakpoints and denser genetic marker data

available for the NILs, we narrowed down the QTL region to about 630 Kbps, containing 22 predicted gene models.

Natural variation in metabolic traits is often caused by *cis* polymorphisms in metabolic enzyme-encoding genes (Meihls *et al.*, 2013; Yan *et al.*, 2015; Handrick *et al.*, 2016). However, we found no predicted acetyltransferase genes within our QTL interval. Moreover, by plotting the distribution of smiglaside C and smilaside A across the IBM RILs, we found an overall positive correlation between these two metabolites (Figure 3.5a), which contradicted the hypothesis of a substrate/product relationship mediated by a polymorphism in a hypothetical acetyltransferase. Closer investigation of the smiglaside C-smilaside A distribution plot revealed that the 83 RILs can be divided by linear discriminant analysis into a B73-like group with lower smilaside A/smiglaside C ratios, and a Mo17-like group with higher ratios (Figure 3.5a). We hypothesized that this phenotypic difference could be attributed to transcriptional regulation,

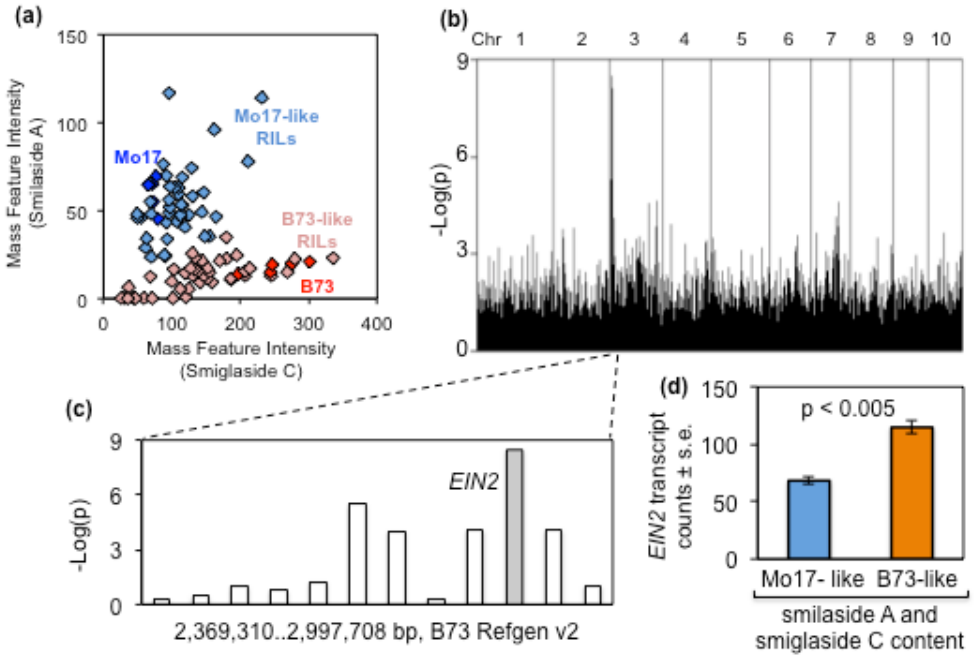


Figure 3.5. *ETHYLENE INSENSITIVE 2* is differentially expressed in B73 × Mo17 recombinant inbred lines (RILs) with contrasting smilaside A/smiglaside C ratios. (a) B73 × Mo17 RILs can be divided into two groups based on their constitutive smilaside A and smiglaside C content. One replicate each of 83 B73 × Mo17 RILs and five replicates of the B73 and Mo17 parental lines were plotted based on the constitutive content of smilaside A and smiglaside C in their seedling roots. The parental lines are indicated with dark blue (Mo17) and dark red (B73). The RILs were determined to be Mo17-like (light blue) or B73-like (light red) in their smilaside A and smiglaside C content using the linear discriminant analysis. The level of significance in differential expression between Mo17- and B73-like inbred lines, measured by -log(p) from Student's *t*-tests, is plotted for each root-expressed transcript in the chromosomal order across the whole genome (b) and within the QTL region on chromosome 3 (c). (d) Expression of *EIN2* is significantly higher in the seedling roots of RILs with a B73-like smilaside A and smiglaside C content than in ones with a Mo17-like content. Mean +/- s.e., P value is from a two-tailed Student's *t*-test.

which would differ between the IBM RILs belonging to either phenotypic group. We therefore performed whole transcriptome profiling on the seedling root samples that were used for metabolite quantification. This whole-genome analysis showed that the genes with the most significant differential expression between the two phenotypic groups are located in the identified QTL region on chromosome 3 (Figure 3.5b,c; Table S3.5). Specifically, the gene showing the most significantly different expression was a positive regulator of ethylene signaling in maize, *ZmEIN2* (*ETHYLENE INSENSITIVE* 2; GRMZM2G068217), which was expressed at a higher level in the seedling roots of RILs with a B73-like abundance of smiglaside C and smilaside A (Figure 3.5d). This leads to the hypothesis that *ZmEIN2*, and hence ethylene signaling, is a negative regulator of smilaside A/smiglaside C ratio.

Acetylated feruloylsucroses accumulation and resistance against *F. graminearum* are regulated by ethylene

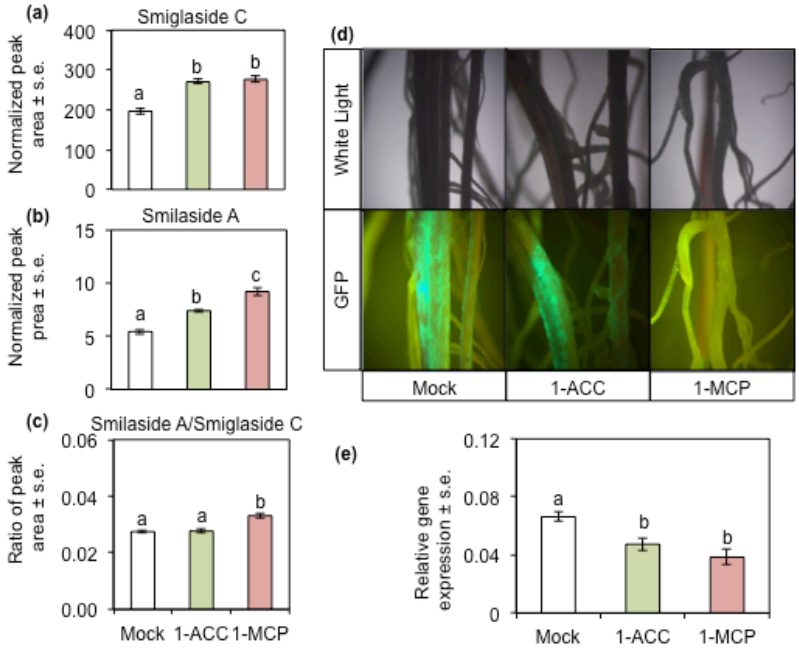


Figure 3.6. Exogenous 1-methylcyclopropene (1-MCP) treatment promotes maize seedling root resistance against *F. graminearum*. B73 maize seedling roots inoculated with *F. graminearum* were treated with 1-aminocyclopropane-1-carboxylate (1-ACC), 1-MCP, or water as a control. The abundance of (a) smiglaside C, (b) smilaside A , and (c) the ratio of the two metabolites was calculated from peak area under negative electron spray ionization mode. Mean +/- s.e. of N = 5, different letters indicate significant difference, $P < 0.05$, ANOVA followed by Tukey's HSD test. (d) B73 maize seedling roots inoculated with *F. graminearum*-GFP and treated with 1-ACC, 1-MCP, or mock treatment for ten days were examined with white light and fluorescence microscopy. More GFP marker expression from *F. graminearum* was observed on root surface of mock- and 1-ACC treated seedling roots than on 1-MCP-treated roots. (e) Fungal growth was quantified by qRT-PCR using *F. graminearum*-specific primers, relative to expression measurement of a maize housekeeping gene. Mean +/- s.e. of N = 8, different letters indicate significant differences, $P < 0.05$, ANOVA followed by Tukey's HSD test.

No *ZmEIN2* mutation is available in public maize transposon insertion collections. Instead, we manipulated ethylene response *in vivo* with a gaseous competitive inhibitor, 1-methylcyclopropene (1-MCP), and a biochemical precursor of ethylene production in plant, 1-

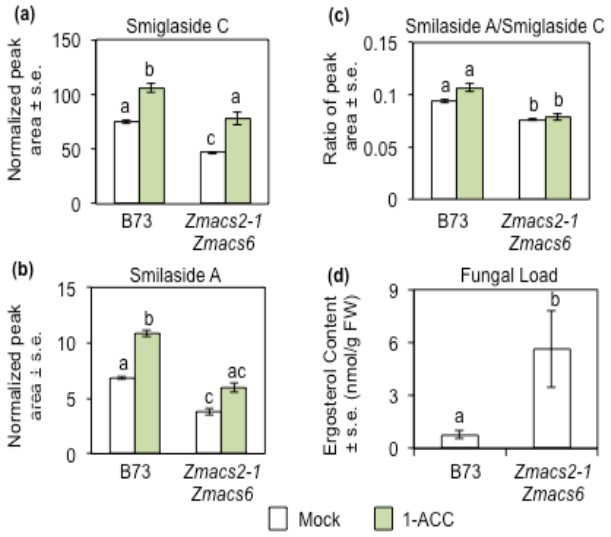


Figure 3.7. Ethylene biosynthesis is required for acetylated feruloylsucrose accumulation and resistance against *F. graminearum*. The abundance of (a) smiglaside C and (b) smilaside A was estimated by peak area at their respective *m/z* ratio under negative electron spray ionization mode. (c) The ratio of smilaside A and smiglaside C peak areas. Mean +/- s.e., different letters indicate significant differences, $P < 0.05$, Two-way ANOVA followed by Tukey's HSD test. (d) Fungal load was estimated by ergosterol content, normalized by root tissue fresh weight. Mean +/- s.e., different letters indicate significant differences, $P < 0.05$, One-way ANOVA.

aminocyclopropane-1-carboxylic acid (1-ACC). Effects of these treatments were confirmed by contrasting seedling growth rate in these groups (Figure S3.7). Consistent with the genetic mapping results, 1-MCP treatment (and hence lower *ZmEIN2* availability) led to hyper-accumulation of smilaside A and an elevated smilaside A/smiglaside C ratio (Figure 3.6a,c). Unexpectedly, smiglaside C abundance was also significantly increased by 1-MCP treatment (Figure 3.6b). Both smilaside A and smiglaside C were induced by 1-ACC treatment, but their ratio was not significantly affected (Figure 3.6a-c).

To further investigate how smilaside A and smiglaside C are regulated by ethylene production, we measured their abundance in the seedling roots of the *Zmacs2-1 Zmacs6* ethylene biosynthetic mutant (Young *et al.*, 2004), which has *Mutator* transposon insertions in two 1-ACC synthase genes in the B73 genetic background. Consistent with prior measurement of lower leaf ethylene content (Young *et al.*, 2004), this double mutant had a lower root ethylene concentration than wildtype (Figure S3.8). Metabolite abundance was measured with and without exogenous 1-ACC, an ethylene biosynthesis intermediate that is downstream of the two mutated *ZmACS* genes. Constitutively, there were significantly lower amounts of both smilaside A and smiglaside C in the roots of *Zmacs2-1 Zmacs6* compared to wildtype B73. After 1-ACC treatment, both metabolites were increased in wildtype B73, and were restored to wildtype levels

in *Zmacs2-1 Zmacs6* (Figure 3.7a,b). In *Zmacs2-1 Zmacs6*, the smilaside A/smiglaside C ratio was significantly lower than in wildtype B73. However, consistent with results from the previous experiment, 1-ACC treatment did not affect this ratio in either genetic background (Figure 3.7c).

To investigate how ethylene production and sensitivity can affect maize seedling defense against *F. graminearum,* we compared *F. graminearum*-inoculated maize seedling roots treated with 1-ACC, 1-MCP, or water control. Although we observed extensive fungal hyphae of GFP-transformed *F. graminearum* on the root surface of both mock and 1-ACC treated seedlings, hyphae were almost completely absent from 1-MCP treated roots (Figure 3.6d). In support of the microscopic observations, we found significantly lower *FgTUB* expression in 1-MCP treated seedling roots (Figure 3.6e). Neither 1-ACC nor 1-MCP affected growth of *F. graminearum* on agar plates (Figure S9), indicating that there is not direct toxic effect. Comparing infected seedling roots of *Zmacs2-1 Zmacs6* and wildtype B73, we found significantly higher *F. graminearum* growth on the mutant line (as measured by ergosterol accumulation; Figure 3.7d).

Maize diferuloylsucroses are induced by multiple fungal pathogens

A mass feature likely representing smiglaside C was identified as a maize acyl sugar that was induced after infection with *Colletotrichum graminicola* (anthracnose leaf blight), though without structural confirmation (Balmer *et al.*, 2013). To determine whether induced production of smilaside A and smiglaside C was a more general maize response to fungal infection, we inoculated seedlings with four additional fungal pathogens, *Aspergillus flavus*, *Rhizopus microspores*, *Fusarium verticilloides*, and *Cochliobolus heterostrophus*. Whereas smilaside A was only induced by *F. verticilloides* and *C. heterostrophus*, all four pathogens significantly induced the accumulation of smiglaside C (Figure 3.8). As in the case of *F. graminearum* infection (Figure 3.3a-c), smiglaside C was

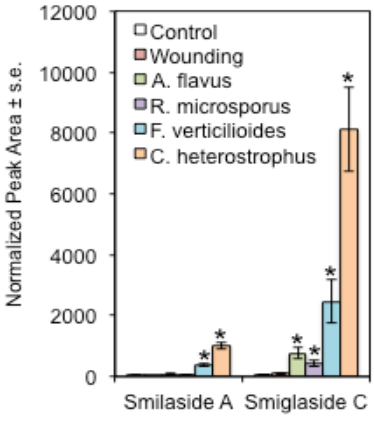


Figure 3.8. Smilaside A and smiglaside C are induced by fungal pathogen infection. Smilaside A and smiglaside C content were measured in 35-day-old Mo17 stem tissues four days after physical wounding or inoculation with *Aspergillus flavus*, *Rhizopus microspores*, *Fusarium verticilloides*, or *Cochliobolus heterostrophus*. Mean +/- s.e. of N = 4. *P < 0.05 relative to the uninfected control, pairwise Wilcoxon rank sum tests with Benjamini and Hochberg correction for multiple comparisons.

induced to a greater extent than smilaside A. Therefore, induced accumulation of smiglaside C may be a general response of maize to infection by fungal pathogens.

Discussion

In addition to maize, acetylated feruloylsucroses may have defensive properties in other plant species. These compounds were first identified in the rhizomes of *S. china* and *S. glabra*, which are used in traditional Chinese medicine (Chen *et al.*, 2000; Kuo *et al.*, 2005). Similar phenylpropanoid sucrose esters, with different numbers and types of phenylpropanoid groups attached, were later found in various Liliaceae and Polygonaceae species. Crude plant extracts containing these compounds, and in some cases purified compounds, have shown anticancer and antioxidant activities *in vitro* (Zhu *et al.*, 2006; Ono *et al.*, 2007; Yan *et al.*, 2008; Zhang *et al.*, 2008; Kim *et al.*, 2010). Building on these promising *in vitro* bioactivities, organic synthesis routes to produce natural phenylpropanoid sucrose esters and structural analogs have been developed with moderate yield and selectivity (Panda *et al.*, 2012a; Panda *et al.*, 2012b). More recently, acetylated feruloylsucroses and other phenylpropanoid sucrose esters were found in rice (Chen *et al.*, 2014; Cho *et al.*, 2015).

Our discovery of smilaside A and smiglaside C in maize will facilitate the *in planta* investigation of phenylpropanoid sucrose ester function and metabolism. Due to the lack of suitable maize mutants, we could not experimentally prove a causal relationship between the genetic polymorphism in *ZmEIN2* and natural variation in constitutive smiglaside C and smilaside A abundance. Instead, we demonstrated that ethylene biosynthesis was required for the production of both compounds *in vivo*, whereas ethylene sensitivity fine-tuned their relative abundance (Figure 6 and 7). Interestingly, fungal load after *F. graminearum* inoculation was negatively correlated with smilaside A/smiglaside C ratio across three natural variation and manipulative experiments in this study (Figure 3, 6, and 7), suggesting that this ratio, rather than the absolute abundance of these two compounds, may be more important for resistance against *F. graminearum*.

Fusarium graminearum infection is known to induce ethylene biosynthetic and responsive genes in both maize seedling roots and *Brachypodium distachyon* spikes (Pasquet *et al.*, 2014). In wheat, different comparative transcriptomic studies have reached opposite conclusions regarding the role of ethylene signaling in responses to *F. graminearum* infection (Li

& Yen, 2008; Ding *et al.*, 2011; Xiao *et al.*, 2013). However, *F. graminearum* resistance could be manipulated by manipulating ethylene signaling in both wheat and barley leaves (Chen *et al.*, 2009). In the current study, we observed that maize seedlings with lower ethylene sensitivity, either due to genetically-encoded polymorphism in *ZmEIN2* expression or artificial treatment with a competitive inhibitor, are more resistant to *F. graminearum* and accumulated the more bioactive smilaside A (Figure 3 and 6). However, exogenous ethylene supplementation in the form of 1-ACC did not promote *F. graminearum* susceptibility in maize seedling roots. This inconsistency could arise from differences in tissue type, developmental stage, or the treatment regime. Together, these results indicate that, above a certain minimal ethylene concentration, ethylene sensitivity negatively regulates the efficiency of biochemical defense, leading to contrasting fungal resistance levels.

In our *in vitro* assay, we observed that the diacetylated smilaside A caused greater fungal growth inhibition than the triacetylated smiglaside C (Figure 3f). This is perhaps surprising, because phenylpropanoid sucrose esters with higher degrees of acetylation have generally shown stronger *in vitro* bioactivities, though these two specific compounds have not been compared previously (Panda *et al.*, 2012b; Cho *et al.*, 2015). Our observations may be explained by a non-linear relationship between the degrees of acetylation and bioactivity. Different structural isomers also may be relevant for the structure-activity-relationship. Other than the putative tetraacetylated diferuloylsucrose, all acetylated diferuloylsucroses detected in our LC-MS analyses showed signs of multiple structural isomers, though there was usually a predominant one (Figure S2). How these different isomers could differ in their bioactivities is another question that requires further investigation.

Since acetylated feruloylsucroses can be induced by various fungal pathogens in maize, it would be interesting to assess their crop protection value *in vivo* under more relevant field conditions. Simple phenylpropanoids can contribute to disease resistance not only through their direct antimicrobial activities, but also by playing a critical role in physical fortification of plant cell walls (Nicholson & Hammerschmidt, 1992). The importance of cell walls as a physical barrier against *F. graminearum* and other fungal pathogens is further highlighted by the prevalence of genes that are likely to encode cell-wall-degrading enzymes in the genomes of fungal phytopathogens (Cuomo *et al.*, 2007; Kubicek *et al.*, 2014). Should acetylated

feruloylsucroses also contribute to the physical strength of plant cell walls, such effects would not be evident in *in vitro* assays.

Although our study has not revealed any enzyme-encoding genes that are directly involved in the biosynthesis of the metabolites of interest, the chemical structures of smilaside A and smiglaside C give us hints about their possible biosynthetic pathway. Specifically, we hypothesize that distinct but related hydroxycinnamoyl transferases are responsible for the esterification of feruloyl-CoA and the free hydroxyl groups on the sucrose molecule, analogous to the phenylpropanoid quinic acid esterification reactions. In the B73 genome, there are 13 predicted hydroxycinnamoyl transferase-encoding genes, mostly with unconfirmed activity and substrates (Schnable *et al.*, 2009). These predicted gene models are candidates for elucidation of the biosynthetic pathway of acetylated feruloylsucroses in maize. Compared to the feruloyl esterification enzymes, the identities of the acetyltransferases that catalyze the acetylation on the glucose ring are less clear. We speculate that these enzymes probably belong to the diverse BAHD acyltransferase family, similar to the acylsugar acyltransferases found in tomatoes (Kim *et al.*, 2012; Schilmiller *et al.*, 2012). Our results lead to the prediction that one or more of these acyltransferases would be positively regulated by ethylene signaling in maize seedling roots. The exact order of feruloyl esterification and acetylation on the sucrose molecule also remains unclear. In rice, non-acetylated 3,6 diferuloylsucrose is detected at a very low level in bulk root extract, suggesting that the acetylation occurs after feruloyl esterification (Cho *et al.*, 2015). However, we did not detect the same compound in our microliter-scale LC-MS analyses of maize roots

Finally, this study demonstrates the feasibility of combining metabolomics, transcriptomics and quantitative genetics methods to elucidate regulation of previously unknown antifungal metabolites in maize. An expansion of this integrative approach to a larger number of maize inbred lines in a genome-wide association study likely will identify both additional metabolites and genes involved in their metabolism. Such genes will be useful in future breeding efforts to enhance the pathogen resistance during maize seedling establishment.

Materials and Methods

Plant growth and fungal inoculation. All maize lines were obtained from the Maize Genetics Cooperation Stock Center (Urbana Champaign, Illinois). Maize seeds were germinated in

moisturized rolls of germination paper, and seedlings were transplanted to 7.5 cm x 7.5 cm plastic pots with Turface ® MVP ® calcined clay (Profile Products LLC, Buffalo Grove, IL) when their primary roots reached approximately 9 cm. For inoculation experiments, seedling roots were immersed in *F. graminearum* spore suspension or mock solution (0.03% Phytagar) for one hour prior to transplanting. Spore suspensions were freshly prepared by flooding and scraping 7-day-old fungal cultures, maintained on potato dextrose agar plates, with 0.03% Phytagar suspension. The spore concentration was adjusted to be 5 x 10^5 spores per ml after measurement with a hemocytometer and light microscopy. Hyphal fragments also were observed in the spore suspension. In all seedling inoculation experiments, *F. graminearum* strain ZTE, which was obtained from Dr. Frances Trail (Guenther & Trail, 2005), which was derived from a field-collected strain Z-3639 (Proctor *et al.*, 1995) transformed with the plasmid pTEFEGFP (Vanden Wymelenberg *et al.*, 1997), with permission from Dr. Robert Proctor. Transplanted seedlings were maintained in long day (16 hours) growth condition at 26 ºC and 70% relative humidity. Six days after inoculation, the total root length of seedlings was measured with the RootReader 2D system (Famoso *et al.*, 2010). For screening of natural variation in *F. graminearum*-induced morphological changes, at least five seedlings of each plant line were mock- or fungus-inoculated in two independent experiments for root length measurement. In two other experiments, seven B73 and Mo17 seedlings were inoculated as described above, root length was measured, and tissue was harvested for further analysis.

***Fusarium graminearum* gene expression measurement, mycotoxin quantification, and visual symptom scoring.** The expression of the *F. graminearum* β-tubulin gene in seedling roots was measured using a previously published protocol (Lou *et al.*, 2016). The primers used for qRT-PCR detection of *F. graminearum* are FgTub-qF293 (5'-ATGCTTCCAACAACTATGCT-3') and FgTub-qR411 (5'-AACTAGGAAACCCTGGAGAC-3'), which were designed based on the *F. graminearum* strain PH-1 reference genome sequence (Cuomo *et al.*, 2007). Fungal gene expression in each technical replicate was normalized by measurement of the expression of a constitutive maize actin gene with the following primers ZmActin-qF (5'-CCATGAGGCCACGTACAACT-3') and ZmActin-qR (5'-GGTAAAACCCCACTGAGGA-3').

Approximately 100 mg of fresh frozen seedling root tissue of each sample was used for extraction of deoxynivalenol, the main *F. graminearum* mycotoxin. Three microliters of 50:49.9:0.1 methanol:water:formic acid extraction solvent (Sigma-Aldrich) were added per mg

of tissue. Ground tissue and extraction solvent were mixed and incubated at 4 ºC for 40 min. Solid debris was separated from the solvent by centrifuging at 10,000 x *g* for 10 min. For each sample, 200 µL clear extract were filtered through a 0.45 micron filter plate by. Deoxynivalenol content in the extract was then measured with an Enzyme-Linked Immuno-Sorbent Assay (ELISA) kit following the manufacturer's protocol (Helica Biosystems, Santa Ana, CA). Visual symptoms in seedling roots were scored for in three other comparative experiments with B73 and Mo17 seedlings 3 weeks after *F. graminearum* inoculation. In both experiments, at least 10 seedling roots of either genotype were scored on a 0-4 scale (0 = symptom-less; 1 = single restricted necrosis spot; 2 = single extended necrosis or multiple restricted necrosis spots; 3 = widespread necrosis throughout; 4 = seedling dead).

Maize root metabolomics analyses. The extraction protocol for deoxynivalenol also was used for plant specialized metabolite extraction. Root extracts were analyzed using liquid chromatography-mass spectrometry (LC-MS). Chromatography was performed on a Dionex 3000 Ultimate UPLC-diode array detector system coupled to Thermo Q-Exactive mass spectrometer. Root extract samples were separated on a Titan C18 7.5 cm x 2.1 mm x 1.9 µm, Supelco Analytical Column (Sigma-Aldrich) with a flow rate of 0.5 ml/min, using a gradient flow of 0.1 % formic acid in LC-MS grade water (eluent A) and 0.1 formic acid in acetonitrile (eluent B). Initial metabolite profiling experiments involved an 8-minute linear gradient from 95:5 A:B to 0:100 A:B for comparison of *F. graminearum* induced metabolomic changes in B73 and Mo17 seedling roots. This method was extended to a 15-minute gradient for better separation in later experiments. Mass spectral parameters were set as follows: Spray voltage 3500 V, Capillary temperature 300ºC, Sheath Gas 35 units, Auxiliary Gas 10 units, probe heater temp 200ºC with a HESI probe. Full scan mass spectra were collected (R:35000 FWHM at m/z 200; mass range: m/z 100 to 900) in both positive and negative ion spray modes. For non-targeted metabolomics analyses, metabolite abundance was estimated with signal intensity acquired through the XCMS-CAMERA mass scan data processing pipeline (Tautenhahn *et al.*, 2008; Benton *et al.*, 2010; Kuhl *et al.*, 2012). Smilaside A and smiglaside C abundance was estimated using peak areas at the respective *m/z* channel under negative electron spray ionization mode. Metabolite quantification was normalized by the total ion concentration to account for technical variation between samples. For non-targeted comparative metabolomics analyses, Student's *t*-tests were used to identify mass features that were constitutively different between

B73 and Mo17 ($p<0.01$, and fold change >2 or $0.01<p<0.05$ and fold change >4) and significantly changed by *F. graminearum* in B73 ($p<0.01$, and fold change >1.5 or $0.01<p<0.05$ and fold change >4).

Structural identification of smiglaside C and smilaside A. To determine the chemical structures of smiglaside C and smilaside A, the LC-MS method for non-targeted metabolomics analysis was adapted to extract bulk maize seedling roots with 100% methanol at 4 ºC overnight. Solid debris was filtered out and the crude extract was concentrated with a Buchi Rotovapor. The concentrated crude extract was fractionated on a normal phase column with a methanol:dichloromethane gradient on a CombiFlash Rf+ (Teledyne Isco, Lincoln, NE), and further purified for the target compounds with a water:acetonitrile gradient on a ZORBAX Eclipse XDB C18 column on an Agilent 1100 HPLC system (Agilent, Santa Clara, CA). Purified compounds were dried, weighed, and re-dissolved in methanol. NMR spectroscopy analyses were carried out on a Unity INOVA 600 instrument (Varian Medical Systems, Palo Alto, CA) with the following conditions: 256 scan for 1H NMR; NT = 16 and NI = 800 for COSY and HSQC; NT = 32 and NI = 1600.

Determination of *in vitro* antifungal activity of smiglaside C and smilaside A. By running a standard curve with the purified compound using the same LC-MS method with incremental injection volume, the concentrations of smiglaside C and smilaside A were determined to be approximately 0.2 mM and 0.1 mM in *F. graminearum*-induced Mo17 seedling roots, respectively. Purified compounds were re-dissolved in dimethyl sulfoxide (DMSO) to ten-fold of their respective *in vivo* concentrations (*i.e.* 2 mM smiglaside C, and 1 mM smilaside A). An *F. graminearum* spore and hyphae suspension was prepared as described above, 80 uL of it was mixed with 100 uL potato dextrose broth and 20 uL of the testing compounds in a 96-well plate. Twenty uL of pure DMSO were included as the negative control for this experiment. Fungal growth was monitored by light absorbance measurement at 405 nanometers at 30-minute intervals for 12 hours at 28 °C. All treatment groups were measured in at least four replicates for statistical comparison.

QTL mapping of the constitutive content of smiglaside C, smilaside A, and their ratio. For genetic mapping of constitutive metabolite abundance, three seedlings were germinated for each Intermated B73 × Mo17 (IBM) recombinant inbred line (RIL) and five for each of the parental lines, B73 and Mo17. Ten-day-old root tissues from the three seedlings of each RIL were pooled

into one sample for LC-MS analysis, whereas the five B73 and Mo17 seedlings were analyzed individually to allow comparison of their constitutive metabolomes. QTL mapping analysis was performed based on published genotype data using the composite interval mapping algorithm implemented in WinQTL Cartographer v2.5 (Wang, 2012).The significance thresholds of these QTL mapping results were determined with five hundred permutations. For the ratio mapping, the smilaside A/smiglaside C ratio was used.

Maize root transcriptome analysis. To ensure comparability between metabolome and transcriptome datasets, total RNA was extracted with the Promega SV Total RNA Isolation Kit from another aliquot of the same seedling root tissues used for non-targeted metabolite profiling. In total, seedling root RNA samples were obtained for 73 IBM RILs and five replicates each of B73 and Mo17. mRNA sequencing libraries were prepared robotically on a Biomek NXp with manual post-PCR cleanup using the Lexogen QuantSeq 3′ mRNA-Seq Library Prep Kits (Kremling *et al.*, 2018). These libraries were pooled into one lane and sequenced with 90 base pairs single end reads on an Illumina NextSeq 500 with v2 chemistry at the Cornell Biotechnological Resource Center. Raw RNAseq read data were converted to SAM format, and aligned to B73 RefGen_v3 5b+ gene models using the STAR v2.5.1 RNAseq-aligner (Dobin *et al.*, 2013). The raw transcript counts were calculated for each gene model in each sample using the HTseq 0.6.1p2 python module (Anders *et al.*, 2015). Finally, gene models with fewer than 10 raw read counts in any one of the 73 samples were filtered out, and raw transcript count for each gene model was normalized by the total transcript count of each sample.

1-aminocyclopropane-1-carboxylic acid (ACC) and 1-methylcyclopropene (1-MCP) Treatment. *Zmacs2-1 Zmacs6* double mutant seedlings, along with the wildtype B73 progenitor, were grown in Turface (Young *et al.*, 2004). When the seedlings emerged approximately 2 cm above the soil line, they were treated with either 100 uL of 50 mM 1-ACC solution or water control for 5 consecutive days.

In a separate set of experiments, *F. graminearum*-inoculated B73 seedlings were transplanted into Turface and kept in 70 L airtight boxes. The bottom of the boxes were filled with water and EthylBloc (0.14% 1-MCP) sachets to reach a final concentration of 5 g/L 1-MCP. Pots containing maize seedlings were elevated from water surface to avoid direct contact with the solution. 1-ACC and water control treatments were conducted using the same setup to ensure

comparability with the 1-MCP treated seedlings. The predicted plant growth effects of 1-ACC and 1-MCP treatments were confirmed by measuring seedling heights in each treatment group.

To confirm that *F. graminearum* growth was not directly influenced by 1-ACC and 1-MCP treatment, plugs of fungal hyphae were transferred to the center of potato dextrose agar plates placed in the sealed 70 L boxes used for seedling experiments. For 1-MCP treatment, EthylBloc was used at the same concentration as described above. For 1-ACC treatment, 50 mM 1-ACC stock solution in DMSO were added to melted and cooled potato dextrose agar to a final concentration of 50 μM. The same final concentration of DMSO was included in plates used for controls and 1-MCP treatments. Fungus radial growth was measured after four days.

For both 1-ACC and 1-MCP experiments, seedling roots were harvested for targeted metabolic analysis with the LC-MS method described above. To confirm that the *Zmacs2-1 Zmacs6* double mutant seedlings are producing less ethylene in their root tissues, 1 mg samples of ground frozen root tissues were placed in airtight 8 mL glass vials for ethylene collection for 29 hours. One mL samples were injected into an Agilent Technologies 6850 Network GC system to estimate ethylene content. The ethylene peak was identified and quantified by comparing to a standard of known concentration, and normalized for tissue weight. Fungus-inoculated seedling roots were examined under an Olympus SZX-12 stereo-microscope with LP Green filter cube to compare fungus spread semi-quantitatively before frozen for LC-MS analysis. Another aliquot of these root tissues was used for q-RT-PCR quantification of *F. graminearum*-specific gene expression to estimate fungal growth, as described above.

Smiglaside C and smilaside A induction by multiple fungal pathogens. Mo17 stem elicitation assays utilized 35-day-old greenhouse grown plants in 1-L pots. Plants in damage-related treatment groups were slit in the center, spanning both sides of the stem, with a surgical scalpel that was pulled 8 to 10 cm upward to create a parallel longitudinal incision. The damage treatments spanned the upper nodes, internodes, and the most basal portion of unexpanded leaves. *Aspergillus flavus*, *Rhizopus microspores*, *Fusarium verticilloides*, and *Cochliobolus heterostrophus* fungal spore inoculations were conducted with 100 μL of water per plant at a concentration of 1×10^7 spores mL^{-1} (Ding *et al.*, 2017). Damage plus water alone was used for a mock inoculation. Localized areas of control and treated stem tissues were covered clear plastic packing tape to minimize tissue desiccation, and stem tissues were harvested 4 d later from each individual plant.

Maize stem tissues were ground to a fine powder in liquid nitrogen and weighed out in 50 mg aliquots. Smilaside A and smiglaside C were analyzed as described previously (Ding *et al.*, 2017). Negative ionization $[M-H]^-$ mode scans (0.1-atomic mass unit steps, 2.25 cycles s-1) from m/z 100 to 1,000 were acquired. Analyses of smilaside A and smiglaside C peak abundance relied on the native parent $[M-H]^-$ ion m/z 777 and m/z 819, and stable average retention times of 12.76 min and 13.94 min, respectively. Both analytes displayed split peaks and for consistency both peaks where integrated and combined for the final analyses.

Ergosterol assays. For ergosterol measurements, *F. graminearum*-treated seedlings were transplanted into pots of twice-autoclaved TX360 Metro Mix and grown for 3 weeks, at which point roots were harvested into liquid nitrogen and stored at -80°C. Ergosterol was analyzed as described previously (Christensen *et al.*, 2014) with the following modifications: roots were crushed and placed in scintillation vials each with 10 ml of chloroform:methanol (2:1 v/v) (99.8%) followed by incubation in darkness overnight at room temperature. One ml of extract from each vial was syringe-filtered through 0.45 µm cellulose acetate membrane filters, and 50 µl of filtrate was added to 50 µL of 10 µM C13-labeled cholesterol (cholesterol-25,26,27-13C; Sigma) in methanol as internal standard. Ergosterol was quantified using an Ascentis Express C-18 column (3 cm × 2.1 mm, 2.7 µm) connected to an API 3200 LC/MS/MS with atmospheric photochemical ionization. The injection volume was 5 µl and the isocratic mobile phase consisted of acetonitrile at a flow rate of 300 µl/min.

Data analysis. All *t*-tests were performed with the ttest function implemented in Microsoft Excel. ANOVA and Wilcoxon rank sum tests were performed in R. Linear discriminant analysis was performed with the SAS software.

References

Ali ML, Taylor JH, Jie L, Sun G, William M, Kasha KJ, Reid LM, Pauls KP. 2005. Molecular mapping of qtls for resistance to gibberella ear rot, in corn, caused by *fusarium graminearum*. *Genome* 48: 521-533.

Anders S, Pyl PT, Huber W. 2015. Htseq: A python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166-169.

Balmer D, de Papajewski DV, Planchamp C, Glauser G, Mauch-Mani B. 2013. Induced resistance in maize is based on organ-specific defence responses. *Plant Journal* 74: 213-225.

Benson JM, Poland JA, Benson BM, Stromberg EL, Nelson RJ. 2015. Resistance to gray leaf spot of maize: Genetic architecture and mechanisms elucidated through nested association mapping and near-isogenic line analysis. *PLoS Genet* 11.

Benton HP, Want EJ, Ebbels TMD. 2010. Correction of mass calibration gaps in liquid chromatography-mass spectrometry metabolomics data. *Bioinformatics* 26: 2488-2489.

Bollina V, Kumaraswamy GK, Kushalappa AC, Choo TM, Dion Y, Rioux S, Faubert D, Hamzehzarghani H. 2010. Mass spectrometry-based metabolomics application to identify quantitative resistance-related metabolites in barley against fusarium head blight. *Molecular Plant Pathology* 11: 769-782.

Brauner PC, Melchinger AE, Schrag TA, Utz HF, Schipprack W, Kessel B, Ouzunova M, Miedaner T. 2017. Low validation rate of quantitative trait loci for gibberella ear rot resistance in European maize. *Theor Appl Genet* 130: 175-186.

Buckler ES, Gaut BS, McMullen MD. 2006. Molecular and functional diversity of maize. *Curr Opin Plant Biol* 9: 172-176.

Chen T, Li JX, Xu Q. 2000. Phenylpropanoid glycosides from *smilax glabra*. *Phytochemistry* 53: 1051-1055.

Chen W, Gao Y, Xie W, Gong L, Lu K, Wang W, Li Y, Liu X, Zhang H, Dong H, et al. 2014. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat Genet* 46: 714-721.

Chen X, Steed A, Travella S, Keller B, Nicholson P. 2009. *Fusarium graminearum* exploits ethylene signalling to colonize dicotyledonous and monocotyledonous plants. *New Phytologist* 182: 975-983.

Cho JG, Cha BJ, Seo WD, Jeong RH, Shrestha S, Kim JY, Kang HC, Baek NI. 2015. Feruloyl sucrose esters from *oryza sativa* roots and their tyrosinase inhibition activity. *Chemistry of Natural Compounds* 51: 1094-1098.

Christensen SA, Nemchenko A, Park YS, Borrego E, Huang PC, Schmelz EA, Kunze S, Feussner I, Yalpani N, Meeley R, et al. 2014. The novel monocot-specific 9-lipoxygenase zmlox12 is required to mount an effective jasmonate-mediated defense against *fusarium verticillioides* in maize. *Mol Plant Microbe Interact* 27: 1263-1276.

Cuomo CA, Gueldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma LJ, Baker SE, Rep M, et al. 2007. The *fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* 317: 1400-1402.

Ding LN, Xu HB, Yi HY, Yang LM, Kong ZX, Zhang LX, Xue SL, Jia HY, Ma ZQ. 2011. Resistance to hemi-biotrophic *F. graminearum* infection is associated with coordinated and ordered expression of diverse defense signaling pathways. *Plos One* 6.

Ding Y, Huffaker A, Kollner TG, Weckwerth P, Robert CAM, Spencer JL, Lipka AE, Schmelz EA. 2017. Selinene volatiles are essential precursors for maize defense promoting fungal pathogen resistance. *Plant Physiology* doi: 10.1104/pp.17.00879. [Epub ahead of print].

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. Star: Ultrafast universal rna-seq aligner. *Bioinformatics* 29: 15-21.

Eichten SR, Foerster JM, de Leon N, Kai Y, Yeh CT, Liu S, Jeddeloh JA, Schnable PS, Kaeppler SM, Springer NM. 2011. B73-mo17 near-isogenic lines demonstrate dispersed structural variation in maize. *Plant Physiology* 156: 1679-1690.

Famoso AN, Clark RT, Shaff JE, Craft E, McCouch SR, Kochian LV. 2010. Development of a novel aluminum tolerance phenotyping platform used for comparisons of cereal aluminum tolerance and investigations into rice aluminum tolerance mechanisms. *Plant Physiology* 153: 1678-1691.

Guenther JC, Trail F. 2005. The development and differentiation of *Gibberella zeae* (anamorph : *Fusarium graminearum*) during colonization of wheat. *Mycologia* 97: 229-237.

Handrick V, Robert CAM, Ahern KR, Zhou SQ, Machado RAR, Maag D, Glauser G, Fernandez-Penny FE, Chandran JN, Rodgers-Melnik E, et al. 2016. Biosynthesis of 8-o-methylated benzoxazinoid defense compounds in maize. *Plant Cell* 28: 1682-1700.

Huffaker A, Kaplan F, Vaughan MM, Dafoe NJ, Ni XZ, Rocca JR, Alborn HT, Teal PEA, Schmelz EA. 2011. Novel acidic sesquiterpenoids constitute a dominant class of pathogen-induced phytoalexins in maize. *Plant Physiology* 156: 2082-2097.

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS, et al. 2017. Improved maize reference genome with single-molecule technologies. *Nature* 546: 524-527.

Kazan K, Gardiner DM, Manners JM. 2012. On the trail of a cereal killer: Recent advances in *fusarium graminearum* pathogenomics and host resistance. *Molecular Plant Pathology* 13: 399-413.

Kebede AZ, Woldemariam T, Reid LM, Harris LJ. 2016. Quantitative trait loci mapping for gibberella ear rot resistance and associated agronomic traits using genotyping-by-sequencing in maize. *Theor Appl Genet* 129: 17-29.

Kim J, Kang K, Gonzales-Vigil E, Shi F, Jones AD, Barry CS, Last RL. 2012. Striking natural diversity in glandular trichome acylsugar composition is shaped by variation at the acyltransferase2 locus in the wild tomato *solanum habrochaites*. *Plant Physiology* 160: 1854-1870.

Kim KH, Chang SW, Lee KR. 2010. Feruloyl sucrose derivatives from *bistorta manshuriensis*. *Canadian Journal of Chemistry-Revue Canadienne De Chimie* 88: 519-523.

Kremling KAG, Chen SY, Su MH, Lepak NK, Romay MC, Swarts KL, Lu F, Lorant A, Bradbury PJ, Buckler ES. 2018. Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature* 555: 520-523.

Kubicek CP, Starr TL, Glass NL. 2014. Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annual Review of Phytopathology* 52: 427-451.

Kuhl C, Tautenhahn R, Bottcher C, Larson TR, Neumann S. 2012. Camera: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Analytical Chemistry* 84: 283-289.

Kuo YH, Hsu YW, Liaw CC, Lee JK, Huang HC, Kuo LMY. 2005. Cytotoxic phenylpropanoid glycosides from the stems of *smilax china*. *Journal of Natural Products* 68: 1475-1478.

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A. 2002. Expanding the genetic map of maize with the intermated b73 x mo17 (ibm) population. *Plant Mol Biol* 48: 453-461.

Li GL, Yen Y. 2008. Jasmonate and ethylene signaling pathway may mediate fusarium head blight resistance in wheat. *Crop Science* 48: 1888-1896.

Lou YR, Bor M, Yan J, Preuss AS, Jander G. 2016. Arabidopsis nata1 acetylates putrescine and decreases defense-related hydrogen peroxide accumulation. *Plant Physiology* 171: 1443-1455.

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, et al. 2009. Genetic properties of the maize nested association mapping population. *Science* 325: 737-740.

Meihls LN, Handrick V, Glauser G, Barbier H, Kaur H, Haribal MM, Lipka AE, Gershenzon J, Buckler ES, Erb M, et al. 2013. Natural variation in maize aphid resistance is associated with 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one glucoside methyltransferase activity. *Plant Cell* 25: 2341-2355.

Mideros SX, Windham GL, Williams WP, Nelson RJ. 2012. Tissue-specific components of resistance to aspergillus ear rot of maize. *Phytopathology* 102: 787-793.

Mueller D (2016a) Corn disease loss estimates from the United States and Ontario, Canada - 2014. Purdue University Extension, Retrieved from http://cropprotectionnetwork.org/crop-loss-estimates/corn-disease-loss-estimates-2014/.

Mueller D (2016b) Corn disease loss estimates from the United States and Ontario, Canada - 2015. Purdue University Extension, Retrieved from http://cropprotectionnetwork.org/crop-loss-estimates/corn-disease-loss-estimates-2015/.

Mueller D (2017) Corn disease loss estimates from the United States and Ontario, Canada - 2016. Purdue University Extension, Retrieved from http://cropprotectionnetwork.org/crop-loss-estimates/corn-disease-loss-estimates-2016/.

Munkvold GP, White DG. 2016. *Compendium of corn diseases*: The American Phytopathological Society. Minneaplois, MN

Nicholson RL, Hammerschmidt R. 1992. Phenolic-compounds and their role in disease resistance. *Annual Review of Phytopathology* 30: 369-389.

Olukolu BA, Wang GF, Vontimitta V, Venkata BP, Marla S, Ji J, Gachomo E, Chu K, Negeri A, Benson J, et al. 2014. A genome-wide association study of the maize hypersensitive defense response identifies genes that cluster in related pathways. *PLoS Genet* 10: e1004562.

Ono M, Takamura C, Sugita F, Masuoka C, Yoshimitsu H, Ikeda T, Nohara T. 2007. Two new steroid glycosides and a new sesquiterpenoid glycoside from the underground parts of *trillium kamtschaticum*. *Chem Pharm Bull (Tokyo)* 55: 551-556.

Panda P, Appalashetti M, Natarajan M, Chan-Park MB, Venkatraman SS, Judeh ZM. 2012a. Synthesis and antitumor activity of lapathoside d and its analogs. *Eur J Med Chem* 53: 1-12.

Panda P, Appalashetti M, Natarajan M, Mary CP, Venkatraman SS, Judeh ZM. 2012b. Synthesis and antiproliferative activity of helonioside a, 3',4',6'-tri-o-feruloylsucrose, lapathoside c and their analogs. *Eur J Med Chem* 58: 418-430.

Pasquet JC, Chaouch S, Macadre C, Balzergue S, Huguet S, Martin-Magniette ML, Bellvert F, Deguercy X, Thareau V, Heintz D, et al. 2014. Differential gene expression and metabolomic analyses of *brachypodium distachyon* infected by deoxynivalenol producing and non-producing strains of *fusarium graminearum*. *Bmc Genomics* 15.

Ponts N, Pinson-Gadais L, Boutigny AL, Barreau C, Richard-Forget F. 2011. Cinnamic-derived acids significantly affect *fusarium graminearum* growth and *in vitro* synthesis of type b trichothecenes. *Phytopathology* 101: 929-934.

Proctor RH, Hohn TM, Mccormick SP. 1995. Reduced virulence of *gibberella zeae* caused by disruption of a trichothecene toxin biosynthetic gene. *Molecular Plant-Microbe Interactions* 8: 593-601.

Schilmiller AL, Charbonneau AL, Last RL. 2012. Identification of a bahd acetyltransferase that produces protective acyl sugars in tomato trichomes. *PNAS* 109: 16377-16382.

Schmelz EA, Kaplan F, Huffaker A, Dafoe NJ, Vaughan MM, Ni XZ, Rocca JR, Alborn HT, Teal PE. 2011. Identity, regulation, and activity of inducible diterpenoid phytoalexins in maize. *PNAS* 108: 5455-5460.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. 2009. The b73 maize genome: Complexity, diversity, and dynamics. *Science* 326: 1112-1115.

Tautenhahn R, Bottcher C, Neumann S. 2008. Highly sensitive feature detection for high resolution lc/ms. *Bmc Bioinformatics* 9.

Vanden Wymelenberg AJ, Cullen D, Spear RN, Schoenike B, Andrews JH. 1997. Expression of green fluorescent protein in *aureobasidium pullulans* and quantification of the fungus on leaf surfaces. *Biotechniques* 23: 686-690.

Wang S, Basten, C.J., Zeng, Z-B. 2012. *Windows qtl cartographer 2.5*. Raleigh, NC: Dept. of Statistics, North Carolina State University.

Xiao J, Jin XH, Jia XP, Wang HY, Cao AZ, Zhao WP, Pei HY, Xue ZK, He LQ, Chen QG, et al. 2013. Transcriptome-based discovery of pathways and genes related to resistance against fusarium head blight in wheat landrace wangshuibai. *Bmc Genomics* 14.

Yan J, Aboshi T, Teraishi M, Strickler SR, Spindel JE, Tung CW, Takata R, Matsumoto F, Maesaka Y, McCouch SR, et al. 2015. The tyrosine aminomutase tam1 is required for beta-tyrosine biosynthesis in rice. *Plant Cell* 27: 1265-1278.

Yan L, Gao W, Zhang Y, Wang Y. 2008. A new phenylpropanoid glycosides from *paris polyphylla* var. Yunnanensis. *Fitoterapia* 79: 306-307.

Yang Q, He Y, Kabahuma M, Chaya T, Kelly A, Borrego E, Bian Y, El Kasmi F, Yang L, Teixeira P, et al. 2017. A gene encoding maize caffeoyl-coa o-methyltransferase confers quantitative resistance to multiple pathogens. *Nat Genet* 49: 1364-1372.

Ye JR, Guo YL, Zhang DF, Zhang N, Wang C, Xu ML. 2013. Cytological and molecular characterization of quantitative trait locus qrfg1, which confers resistance to gibberella stalk rot in maize. *Molecular Plant-Microbe Interactions* 26: 1417-1428.

Young TE, Meeley RB, Gallie DR. 2004. Acc synthase expression regulates leaf performance and drought tolerance in maize. *Plant Journal* 40: 813-825.

Zhang L, Liao CC, Huang HC, Shen YC, Yang LM, Kuo YH. 2008. Antioxidant phenylpropanoid glycosides from *smilax bracteata*. *Phytochemistry* 69: 1398-1404.

Zhang Y, He J, Jia LJ, Yuan TL, Zhang D, Guo Y, Wang YF, Tang WH. 2016. Cellular tracking and gene profiling of *Fusarium graminearum* during maize stalk rot disease development elucidates its strategies in confronting phosphorus limitation in the host apoplast. *Plos Pathogens* 12.

Zhu JJ, Zhang CF, Zhang M, Wang ZT. 2006. Studies on chemical constituents in roots of *rumex dentatus*. *Zhongguo Zhong Yao Za Zhi* 31: 1691-1693.

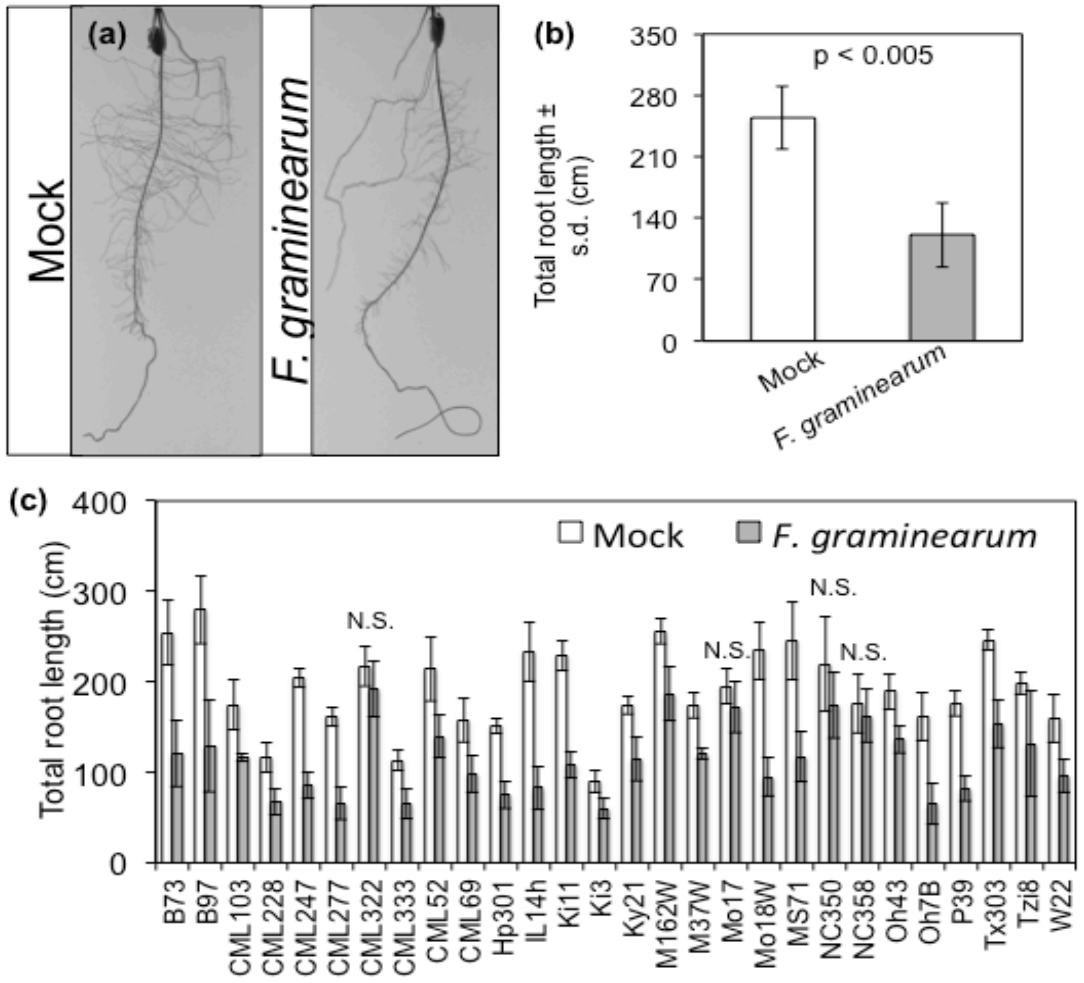


Figure S3.1. Natural variation in *F. graminearum*-induced root growth inhibition. (A) Representative photos of ten-day-old B73 maize seedling roots mock-inoculated or inoculated with a *F. graminearum* spore suspension. Photos were taken at six days post-inoculation. Similar photos were used for total root length measurement. (B) Eight replicates of either mock- or *F. graminearum*-inoculated B73 seedlings were measured to identify differences between the treatment groups, $P < 0.05$, Student's *t*-test. (C) The same assay was expanded to include the 26 parental lines of the nested association mapping population, along with inbred lines Mo17 and W22, to identify natural variation in *F. graminearum*-induced maize seedling root growth reduction. All but four lines (marked by N.S.) showed significant reduction in root growth six days after *F. graminearum* inoculation ($P < 0.05$, two-tailed Student's *t*-test). Root growth ratios based on these data are shown in Figure 1.

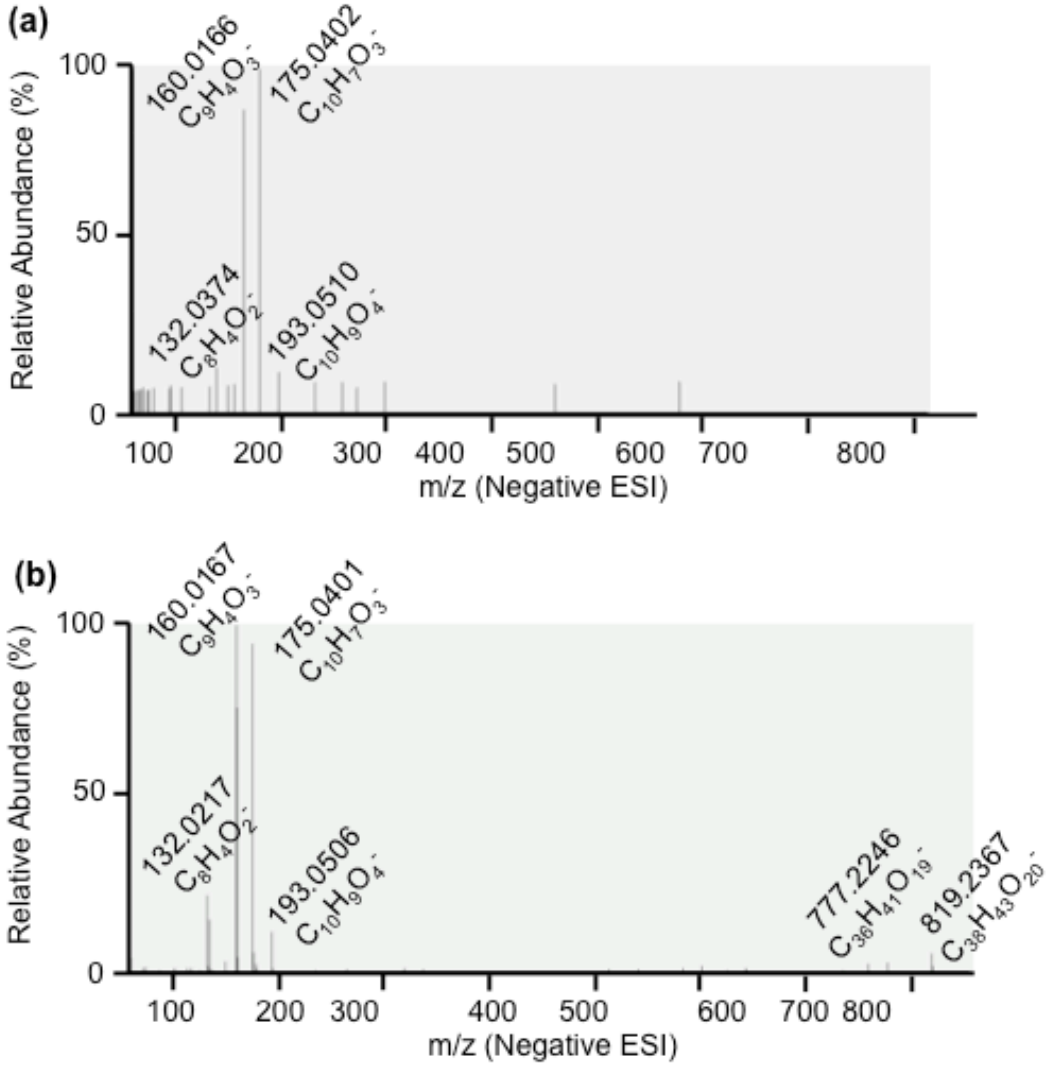


Figure S3.2. Tandem MS spectra of predicted acetylated feruloylsucroses. The tandem MS spectra of the two putative acetylated feruloylsucroses were generated from mock-inoculated Mo17 and B73 seedling root extracts. Scan filters were set to an *m/z* ratio range around their expected parental ion. Major daughter ions are labeled with their exact *m/z* value and most probable predicted molecular formula.

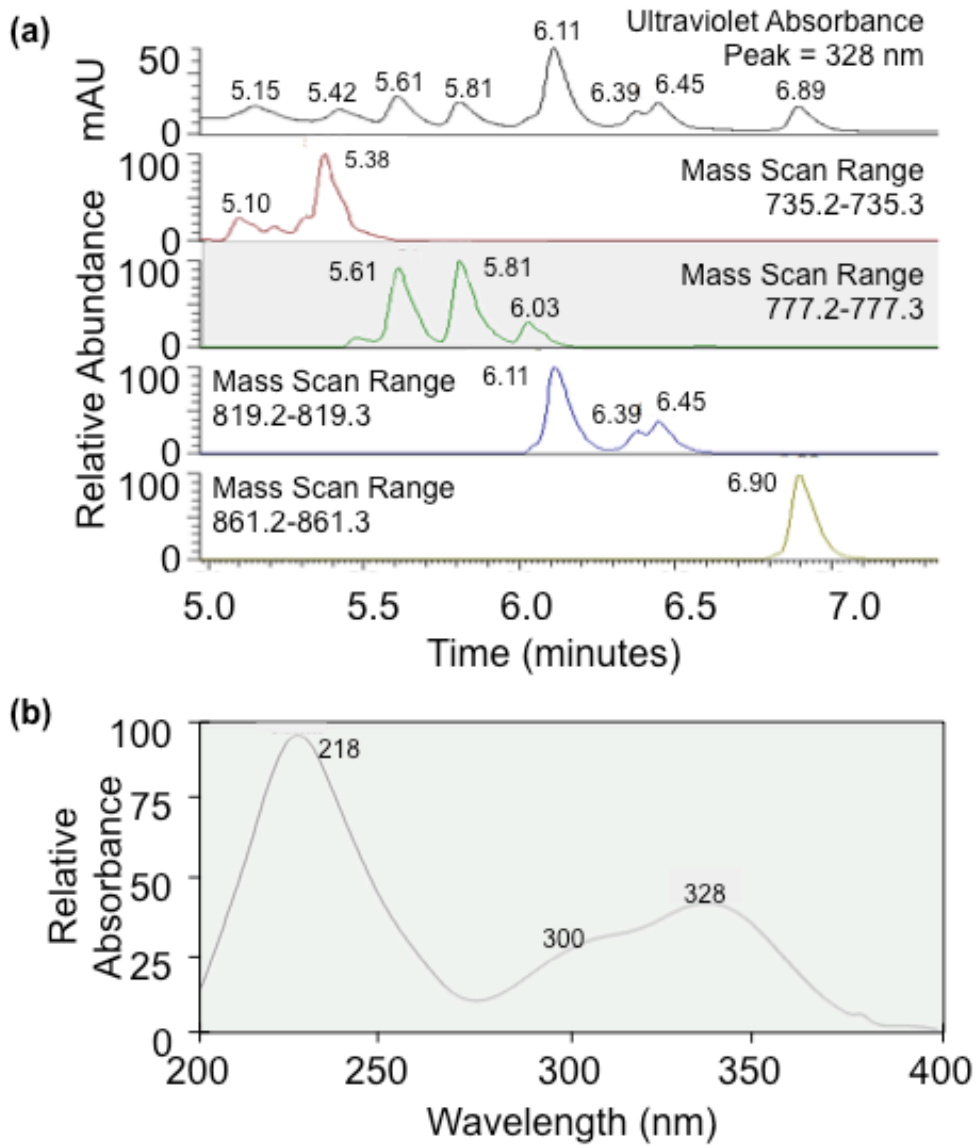


Figure S3.3. Predicted feruloylsucrose compounds have characteristic phenylpropanoid-like ultraviolet absorbance profiles. Feruloylsucrose compounds, as predicted by their exact molecular mass, co-elute with UV absorbance peaks with characteristic phenylpropanoid-like absorbance profiles. UV absorbance data collected from a photo diode array detector are filtered to show only peak absorbance at 328 nm, and mass spectrometry data are filtered for the expected molecular masses of predicted feruloyl acetylsucroses (a). Each peak is labeled with its retention time. A characteristic phenylpropanoid-like UV absorbance profile is shown, with peaks at 218 nm and 328 nm, and a shoulder at 300 nm (b).

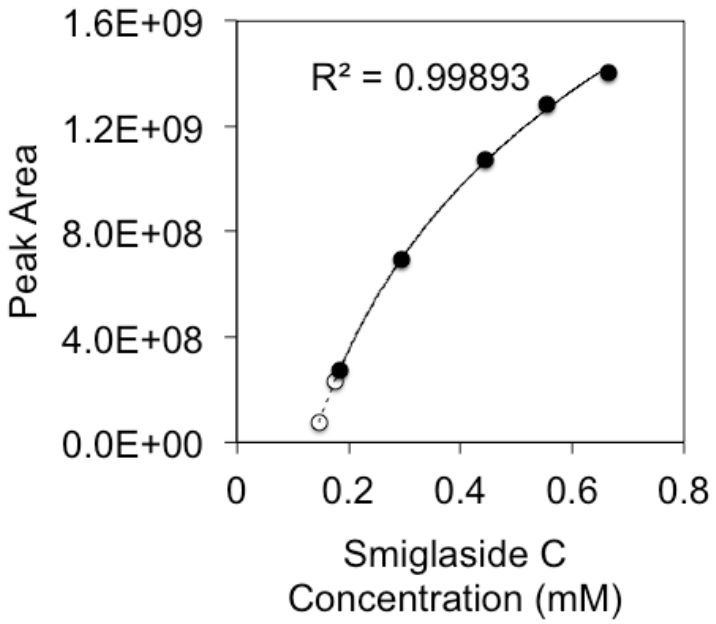


Figure S3.4. Standard curve of smiglaside C. Calculated smiglaside C concentrations from five different injection volumes of the same purified compound solution are plotted with their corresponding mass feature peak areas in filled marks. A logarithmic curve was fitted to these data points with a calculated regression co-efficient (R^2). The same curve is extended to calculate the *in vivo* concentrations of smiglaside C (0.1750 mM) and smilaside A (0.1470 mM) based on MS peak areas in an *F. graminearum*-induced Mo17 seedling root extract. These two peak areas from biological samples and their inferred concentrations are shown in empty circles.

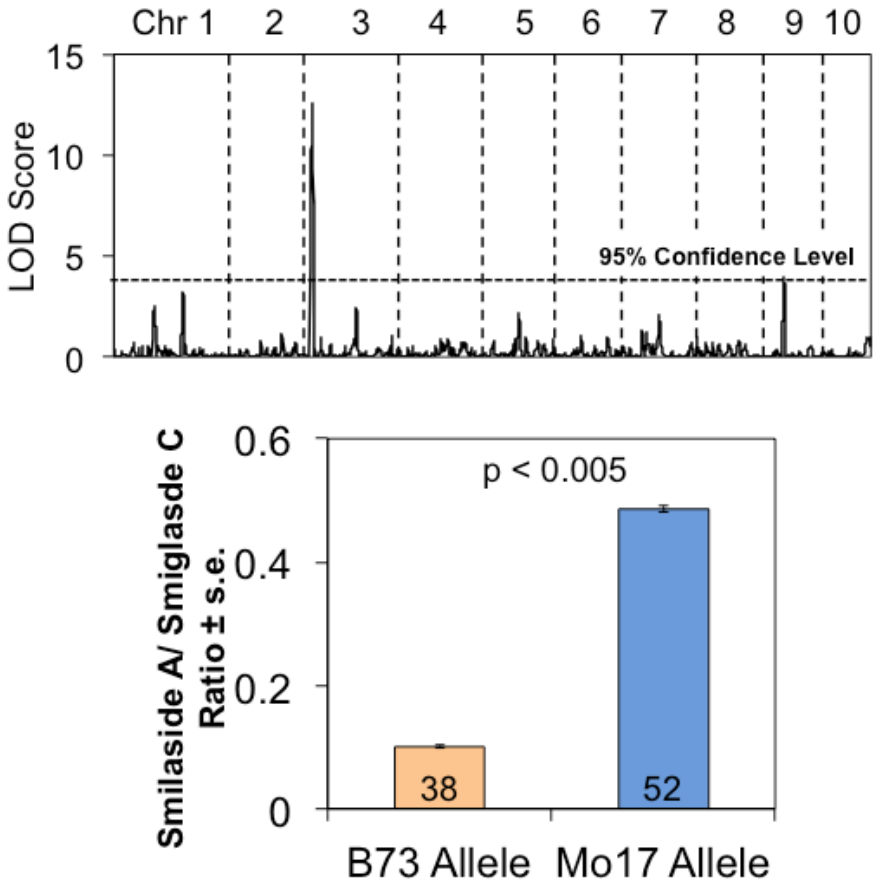


Figure S3.5. QTL mapping of the smilaside A/smiglaside C ratio identifies a locus on chromosome 3. The smilaside A/smiglaside C ratio in the seedling roots of 83 B73 × Mo17 recombinant inbred lines (RILs) was used for composite interval mapping. The calculated log of odds (LOD) score (top panel) of each locus is plotted across the ten maize chromosomes. The 95% confidence interval of the LOD scores was calculated with 500 permutations. RILs and parental lines carrying the Mo17 allele at the most significant QTL on chromosome 3 have a significantly higher smilaside A/smiglaside C ratio than those carrying the B73 allele (lower panel). P-values were calculated using Student's t-test, and the number of RILs in either group is indicated each column.

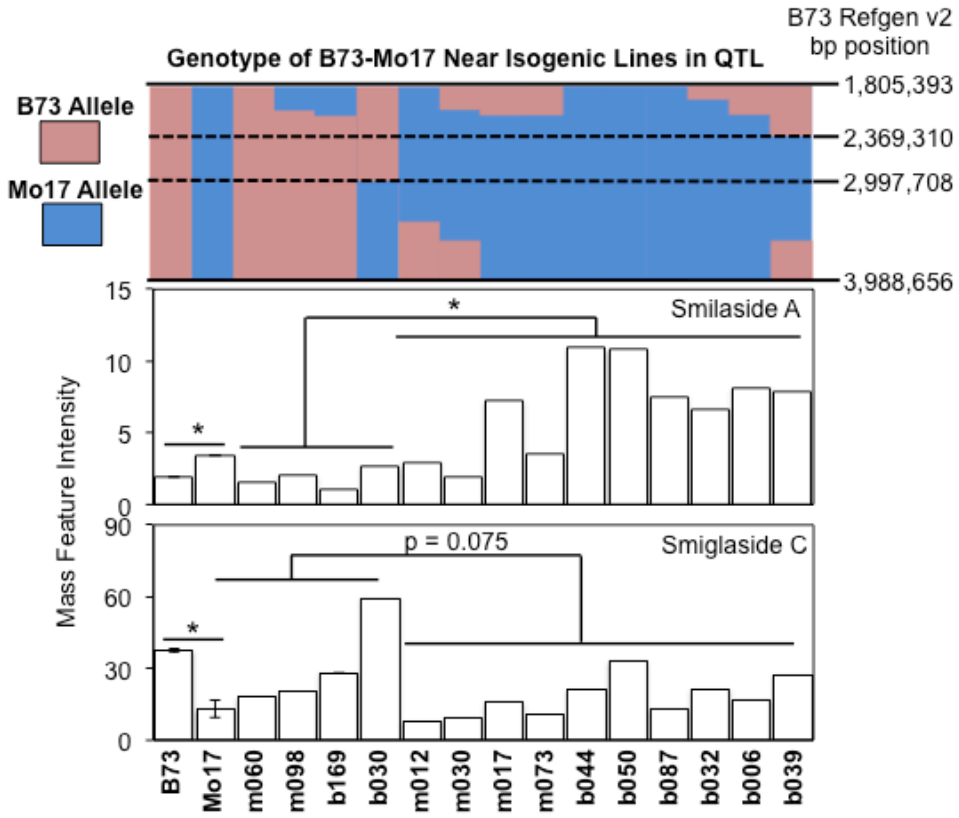


Figure S3.6. Constitutive abundance of smilaside A and smiglaside C varies across B73 × Mo17 near isogenic lines (NILs). The genetic region mapped in the top panel is determined by the 95% confidence interval of a chromosome 3 QTL for smilaside A and smiglaside C abundance. The genetic map of this region and the allelic state of each near isogenic line are reproduced based on data published in Eichten et al., 2011. The genetic background of the NILs is indicated by the initial letter of the line name, *e.g.* m060 has a Mo17 genetic background and b169 has a B73 genetic background Constitutive abundance of smilaside A and smiglaside C were estimated from the mass feature intensity measured in roots of B73, Mo17, and NILs. Mean +/- s.e. of N = 5, *P < 0.05, Student's *t*-test.

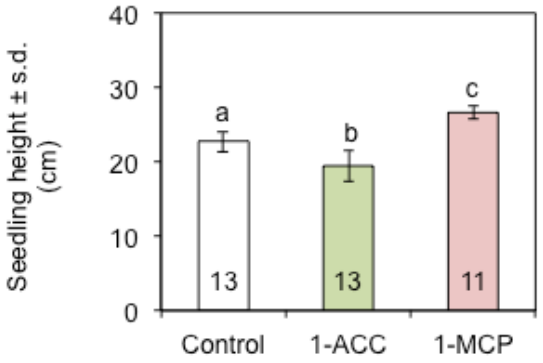


Figure S3.7. 1-Aminocyclopropane-1-carboxylic acid and 1-monocyclopropene treatments have opposite effects on maize seedling height. Seedling heights between each pair of treatment groups are compared to control with one-way ANOVA and Tukey's HSD test, and significant differences ($P < 0.05$) are indicated by different letters. The number of biological replicates is denoted by numbers in each bar.

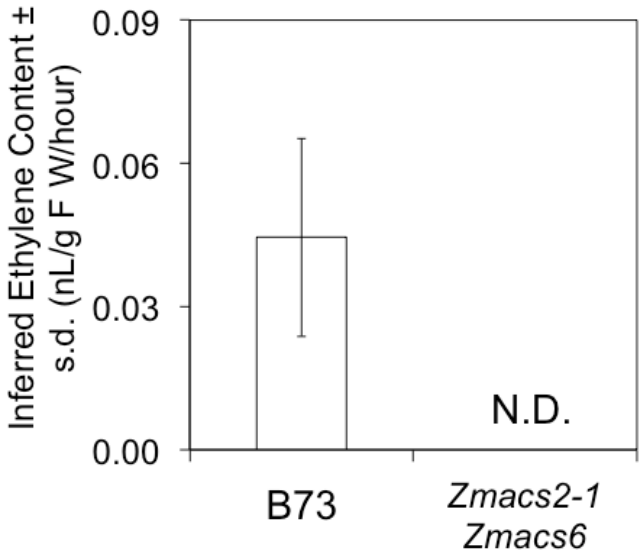


Figure S3.8. Endogenous ethylene production in root is depleted in the *Zmacs2-1 Zmacs6* double mutant maize seedlings. Ethylene produced by root tissues was collected in sealed glass vials for 29 hours and analyzed by gas chromatography. Absolute ethylene content is inferred from peak area measurement from a flame ionization detector. Peaks corresponding to ethylene were not detected (N.D.) in the *Zmacs2-1 Zmacs6* samples. Mean+/- s.e. of N = 6.

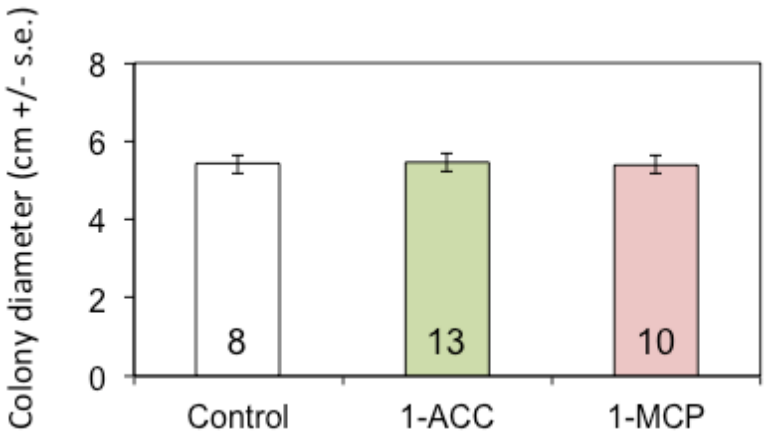


Figure S3.9. 1-Aminocyclopropane-1-carboxylic acid and 1-monocyclopropene are not directly toxic to *F. graminearum.* Fungi were inoculated into the center of potato dextrose agar plates, enclosed in airtight boxes, and colony diameter was measured after five days. 1-ACC was added to the agar at 50 µMconcentration. 1-MCP was released as a volatile using EthylBloc at 5 g/L. Mean +/- s.e., numbers in bars indicate sample sizes, no significant differences $P > 0.05$, two-tailed Student's *t*-tests.

CHAPTER FOUR

A PUTATIVE MAIZE VESICULAR TRANSPORT PROTEIN COULD REGULATE BENZOXAZINOIDS IN MAIZE SEEDLING ROOTS

Abstract

To further take advantage the untargeted seedling root metabolomics dataset collected across the intermated B73 x Mo17 recombinant inbred lines, natural variation in 74 mass features that are significantly different between the two parental lines are used for composite interval mapping. The resulted quantitative trait loci (QTL) are non-uniformly distributed across the maize genome, and a few genomic hotspots emerge, as numerous mass features are significantly associated with them. Within two of these hotspots, the expressions of genes correlate with the mass feature abundance more significantly compared to the rest of the genome. Specifically, a putative vesicular transport protein is negatively associated with benzoxazinoids content through both genetic mapping and transcript-metabolite abundance correlation approaches. Genetic mutant analyses and pharmacological treatment experiments show inconclusive support for this hypothetical association, and hint on the potential regulatory role of vesicular transport on benzoxazinoid metabolism in maize seedling roots.

Introduction

Comparative metabolomics analyses have been a powerful approach to identify specific metabolites associated with biological phenotypes of interest. In plant research, this tool has been used to study resistance against biotic and abiotic stresses as well as to identify biomarkers associated with desirable agronomic traits (Bollina et al. 2010; Pasquet et al. 2014; Ponts et al. 2011; Santiago et al. 2007). More recently, natural variation in specific metabolites of interest has been used to identify novel biosynthetic genes using a quantitative genetics approach (Handrick et al. 2016; Meihls et al. 2013; Richter et al. 2016). Yet, high-throughput metabolomics data are underused when the research focus is placed on a handful of specific metabolites.

In this chapter, we took advantage of the untargeted metabolomics dataset on seedling roots of maize intermated B73-Mo17 recombinant inbred lines generated from the previous experiment, and performed metabolome-scale genetic mapping of over 700 mass features in this population. In result, we found over 26% of these mass features were significantly associated with one or more of six hotspots in the maize genome. We further adopted the transcriptomics dataset generated from the same tissues to pinpoint candidate genes within large QTL intervals using a cross-omics correlation analysis approach. In combination, these methods led us to negatively associate a putative vesicular transport protein, and hence the vesicular transport activities, with benzoxazinoid abundance in maize seedling roots. This hypothesis was partially supported by metabolomics analyses results of genetic mutants and pharmacological treatment experiments.

Results

Genomic hotspots associated with multiple metabolites are revealed by compiling metabolome-scale QTL mapping results.

To gain a system level understanding of maize seedling root metabolome regulation, we used each of the 727 mass features that was significantly different between the two parental inbred lines, B73 and Mo17, as an independent quantitative trait to perform composite interval mapping as described in Chapter Three. Significant QTL for each metabolic trait were then compiled by counting the number of metabolic traits mapped to each genetic locus.

In result, we observed that the significant metabolic QTL were not uniformly distributed across the maize genome, but concentrated towards specific genomic hotspots instead (Figure 4.1). There were 6 genomic hotspots with at least 20 mass features mapped to each of them. They were located on 4 different chromosomes, with physical size varying between 0.7 and 16.8

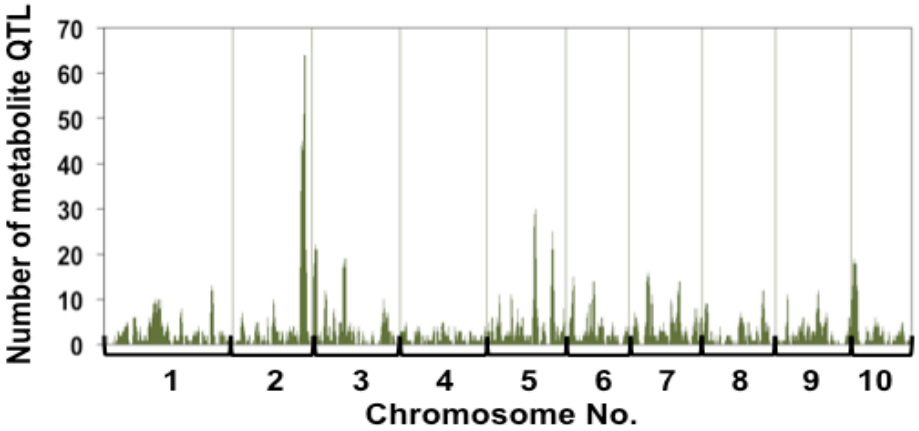


Figure 4.1 Metabolite QTL genomic hotspots revealed by counting co-localized QTL. Genetic loci were laid out on their chromosomal order, and the number of mass features showing a significant QTL at each locus is plotted to demonstrate non-uniform distribution of metabolite QTL across the maize genome.

Mbp, each containing a significant QTL for up to 76 mass features (Table S4.1). Accumulatively, 194 mass features, or 26.7% of all the mass features we mapped, had one or more significant QTL located within one of these six genomic hotspots. Among them, only 33

Table 4.1. Metabolite QTL genomic hotspots from seedling root metabolites in IBM RILs.

Hotspot Number	Number of Mapped Mass Features	Chromosome	Left Marker	Right Marker	Physical Size (Mbp)
1	76	2	umc1252	bnlg1893	7.434
2	27	3	bnl8.15	umc2049	1.418
3	22	3	bnlg1144	bnlg1647	2.993
4	39	5	mmc0081	mmp104	16.833
5	32	5	npi442	bnlg1118	0.762
6	40	10	mmp48a	umc2053	4.685

mass features have a significant QTL in more than one of the six hotspots, suggesting that these genomic hotspots tend to regulate distinct segments of maize seedling root metabolome.

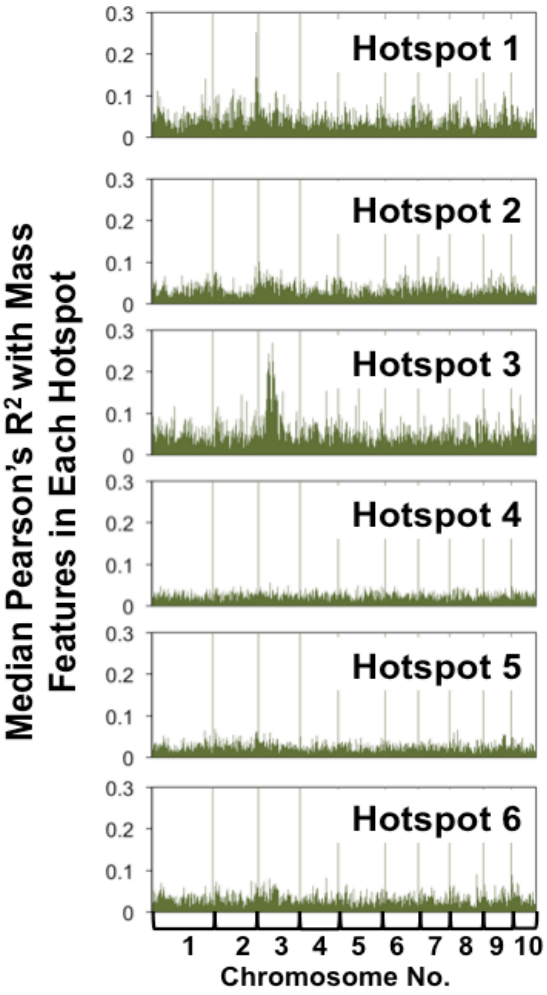


Figure 4.2 Transcripts located in metabolite QTL hotspots sometimes are better correlated with the mass features mapped to the locus than other transcripts. All transcripts mapped to the maize genome are laid out based on their chromosomal order, and their median Pearson's correlation coefficient (R^2) with all mass features mapped to each metabolite QTL hotspot are plotted separately. The numbering of metabolite QTL hotspots are consistent with that in Table 4.1.

Cross-omics correlation analysis can help pinpointing the most probable candidate genes within large QTL intervals.

Limited by genotypic marker density and recombination frequency, genomic hotspots identified from our composite interval mapping analyses can span up to 16.8 Mbp, potentially containing hundreds of genes and numerous genome structure variations (Table 4.1). To pinpoint the causal polymorphism within the large QTL intervals, we hypothesize that the influences of the causal polymorphism on the metabolic traits are mediated by transcriptional regulation. This hypothesis leads to the prediction that the expression of the causal gene within a QTL interval should correlate with the natural variation with the metabolic traits that map to this interval.

To test this hypothesis, we used the transcriptomics data generated from another aliquot of the same seedling root tissues to calculated pairwise Pearson's correlation between each of the 17,000 genes expressed in these tissues and the mass features mapped to each of the six genomic hotspots. To visualize these results, we plotted the median Pearson's correlation coefficient (R^2) between each gene and all the mass features mapped to a genomic hotspot according to the physical genomic location of the gene. As a result, in two of the six genomic hotspots, we observed clear elevation in median R^2 for genes located within the metabolite QTL intervals, whereas this index is uniformly low for the other four hotspots (Figure 4.2). This suggests that some genetic polymorphisms associated with metabolic traits exert their influences

through transcriptional regulation, while others are probably mediated through non-transcriptional mechanisms.

Expression of a putative maize vesicular transport protein is significantly correlated with natural variation in benzoxazinoid compounds.

Within the two genomic hotspots where expressions of contained genes are more significantly correlated with natural variation in mass features mapped to the region than background level, more genes in the hotspot on Chromosome 3 showed elevated correlation coefficients (Figure

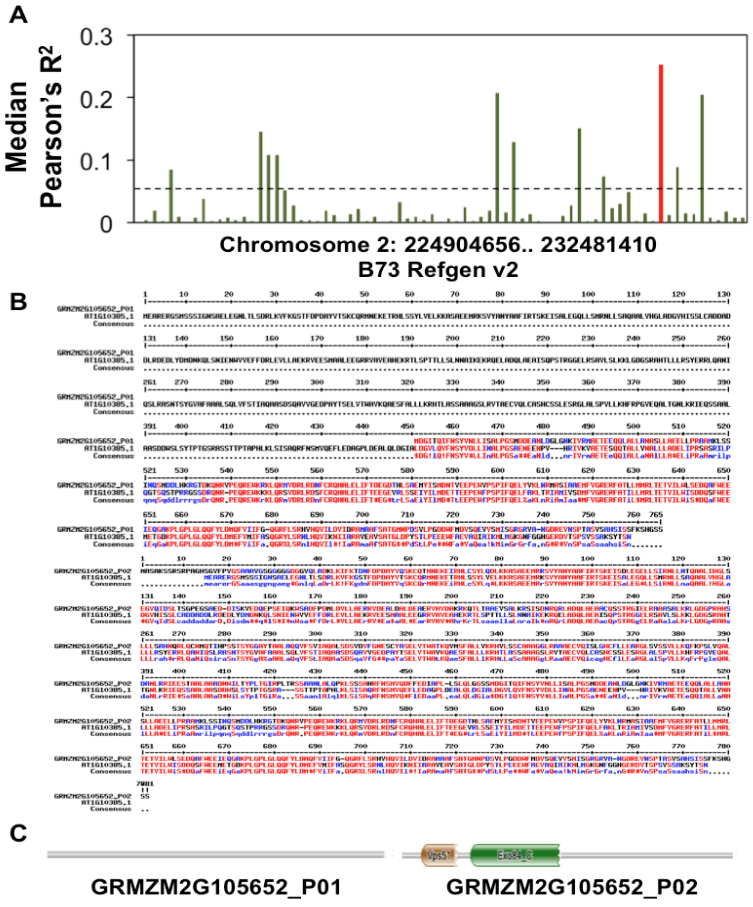


Figure 4.3 The transcript most strongly correlated with mass features mapped to hotspot 2 encodes a putative vesicular transport protein. Transcripts physically located within metabolite QTL hotspot 2 are laid out in their chromosomal order, and their median Pearson's correlation coefficient (R^2) with all mass features mapped to this locus are plotted, showing that GRMZM2G105652 (marked in red) is the most strongly correlated transcript (A). The translated amino acid sequences of the two predicted transcripts of GRMZM2G105652 (top rows) are aligned to their respective closest *Arabidopsis thaliana* orthologs in middle rows. Letters in red indicate fully conserved amino acid residuals, in blue suggest amino acids of similar chemical properties, and un-aligned amino acids are shown in black (B). Both amino acid sequences are also annotated for conserved protein domains with hidden Markov Models (C). In both sequence analysis panels, the left panel shows results from the short predicted protein (GRMZM2G105652_P01), and the right panel shows the longer one (GRMZM2G105652_P02).

4.2). Therefore, while the mass feature-transcript correlation approach led us to the same genomic region that's associated with a group of mass features as the quantitative genetics approach, it did not further enhance the resolution, or help us prioritize any specific candidate gene within the QTL interval.

In contrast, when we zoomed into the genomic hotspot on Chromosome 2, only 11 out of the 75 expressed genes had outstanding median R^2 ($Z > 1.96$) with the mass features mapped to

this region (Figure 4.3A). Among them, the gene model with the highest median R^2 was GRMZM2G105652. Specifically, we found that 62 of the 76 mass features, including many known to be associated with four benzoxazinoid compounds, were negatively correlated with the expression of this gene, and only 14 mass features were positively correlated (Table S4.2). This evidence suggests that this gene is a negative regulator of benzoxazinoid accumulation in maize seedling roots.

GRMZM2G105652 was predicted to encode two different splice forms differing by the presence/absence of one exon in the middle of the transcript. In *Arabidopsis thaliana*, the closest ortholog of the short transcript encoded a Vps51/Vps67 (components of vesicular transport) protein (AT1G10385.1), and the long transcript was orthologous to the exocyst complex component 84B (AT5G49830.1). However, only the long transcript has a full-length alignment with its *Arabidopsis* ortholog, whereas the short transcript only covers the 3' end of the *AtVPS51/VPS67* gene (Figure 4.3B). Consistently, the long transcript was annotated with a Vps51 domain and an exocyst complex component 84 C-terminus domain when analyzed with hidden Markov Model scan, but the same analysis of the short transcript did not identify any known protein domain (Figure 4.3C). These evidences hint that short splice form may be an artifact of the computational annotation pipeline, and indeed, the mRNA sequence record related to this splice form (NM_001152600.1) has been removed from the National Center for Biotechnology Information (NCBI) online database due to insufficient experimental evidence.

Transposon insertions in the putative vesicular transport protein may promote accumulation of benzoxazinoid compounds.

To validate the hypothesized negative regulatory function of the putative vesicular transport protein on benzoxazinoid compounds, we obtained four different Uniform-*Mutator* (UFMu) lines with transposon insertions in the 5' un-translated region (UTR) of the predicted gene model from the Maize Genetics Co-op Center. Propagation of these mutant lines turned out to be difficult, and by the completion time of this thesis, we were able to propagate only two of the four UFMu lines from two heterozygous individuals. Genotyping of their progenies suggests that the homozygous mutants may not be viable since we have recovered no homozygote from 21 seeds planted for either mutant line with 71% and 76% germination rates.

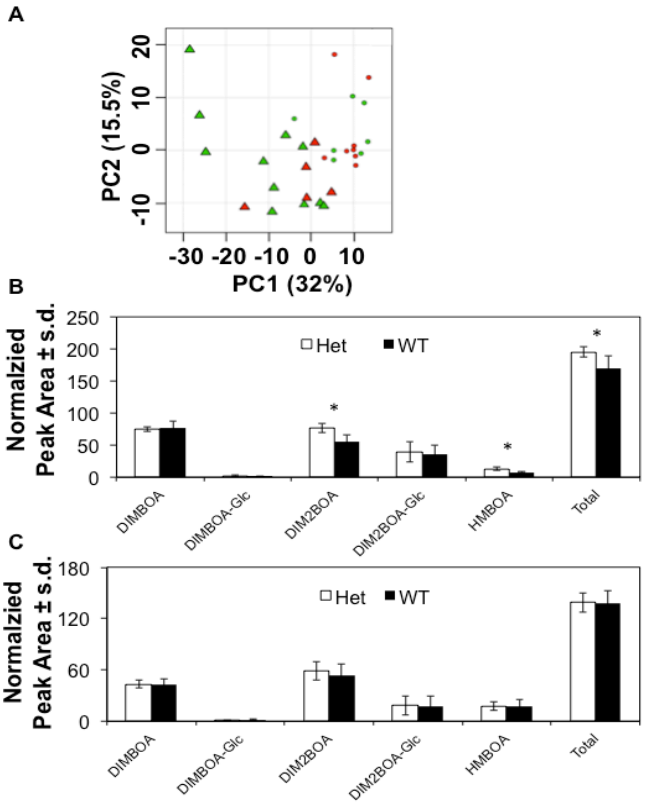


Figure 4.4 Two independent transposon insertion mutant lines demonstrate inconsistent biochemical phenotypes. Metabolome-level differentiation is stronger between the two mutant backgrounds (denoted by different marker shapes) than the presence/absence of target transposon insertion (denoted by different colors, A). Concentrations of specific benzoxazinoid compounds and total benzoxazinoid content in seedling roots are significantly higher in individuals carrying one (Het) versus no (WT) target transposon insertion allele in one genotypic background (B) but this pattern is not found the other genotypic background (C).

We proceeded with non-targeted comparative metabolomic analyses of the seedling root tissues of heterozygous and wildtype individuals with either *Mu* insertions. Principal component analysis of the total metabolome of these individuals revealed clear differentiation between the two UFMu lines, suggesting a strong influence of the background *Mu* insertions scattered elsewhere across the genome (Figure 4.4). However, we did not observe any significant differentiation between wildtype and heterozygous individuals in either mutant genotype background at the metabolome level (Figure 4.4A).

Since many of the mass features significantly correlated with GRMZM2G105652 were representative of three benzoxazinoid compounds, we performed targeted metabolite quantification of these benzoxazinoid compounds in the heterozygotes and wildtypes. One of these compounds, M2BOA, was probably a degradation product of 8-*O*-methylated benzoxazinoids, and was not detected in fresh root extract. In one of the two UFMu lines, we found the heterozygotes accumulated significantly more 2-2,4-dihydroxy-7,8-dimethoxy-1,4-benzoxazin-3-one (DIM_2BOA). Interestingly, the inactive glucoside of DIM_2BOA (DIM_2BOA-Glc), and their biochemical precursors 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA) and DIMBOA-glucoside were not significantly different between the same two groups of individuals, suggesting that the difference in DIM_2BOA abundance is not due to differential activity in its recently identified biosynthetic

enzyme BX13. In the same UFMu line, heterozygotes also accumulated significantly more 2-hydroxy-7-methoxy-1,4-benzoxazin-3-one (HMBOA) than wildtype individuals in their roots, and when we summed up the quantification of all five major benzoxazinoid compounds, we found that the heterozygotes have higher overall benzoxazinoid abundance (Figure 4.4B). However, in the other UFMu line, we observed no significant difference in neither specific nor total benzoxazinoid content between heterozygous and wildtype individuals (Figure 4.4C).

Pharmacological Disruption of Vesicular Transport Induces Accumulation of Benzoxazinoid Compounds in Maize Seedling Roots.

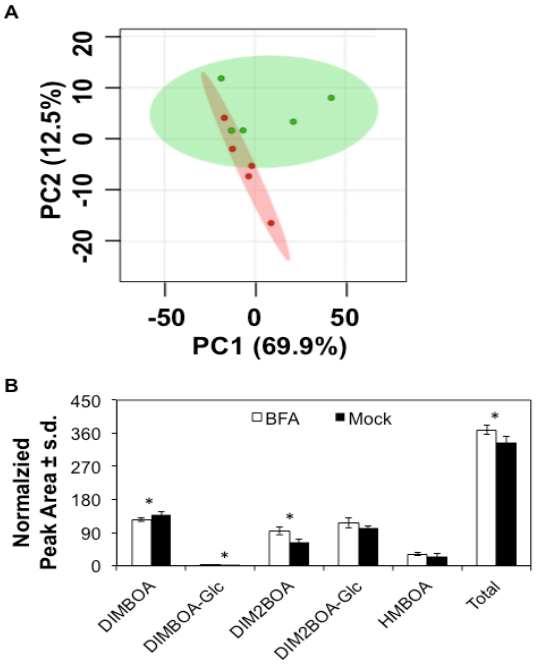


Figure 4.5 Brefeldin A treatment induces significant changes in benzoxazinoid compounds in maize seedling roots. Brefeldin A and mock-treated Mo17 seedling roots can be differentiated at metabolome level **(A)**. Brefledin A-treated seedling roots have significantly higher amount of total benzoxazinoids than mock-treated controls, but the differences are variable for specific benzoxazinoid compounds **(B)**.

GRMZM2G105652 encodes a putative maize exocyst complex component protein, which is required for the exocytosis process (Zarsky et al. 2013). Hence, we hypothesize that the activity of secretory pathway and vesicular transport in maize seedling roots negatively regulate benzoxazinoids accumulation, based on the negative correlation between gene expression and metabolite abundance observed (Table S4.2). Since results from the genetic mutant analyses were inconclusive, we also tested this hypothesis by treating Mo17 seedling roots with a vesicular transport inhibitor, Brefeldin A (BFA). In result, we found clear differentiation between mock- and BFA-treated Mo17 seedling root samples at metabolome level based on PCA (Figure 4.5A). When comparing specific benzoxazinoid compounds BFA treated seedling roots accumulated significantly higher amount of DIMBOA-Glc and DIM_2BOA, but less DIMBOA than mock treated controls, while there were no significant difference in DIM_2BOA-Glc, HMBOA (Figure 4.5B). These results suggest that vesicular transport can indeed regulate benzoxazinoid accumulation in maize seedling roots. However, the specific way of regulation is likely more complicated than we previously hypothesized, since different benzoxazinoid

compounds demonstrate inconsistent patterns in BFA treated seedling roots. Finally, we summed up the quantification of all five benzoxazinoid compounds we examined in each sample, and found that the total benzoxazinoid abundance was higher in the BFA-treated seedling roots, suggesting that vesicular transport overall is a negative regulator of benzoxazinoid accumulation.

Discussion

Co-localizing QTL from plant genetics studies have been a powerful indicator of shared regulator of distinct quantitative traits of interests (Mideros et al. 2014). In studying of plant specialized metabolites, this approach has been successfully deployed to identify key biosynthetic genes, as well as their potential ecological functions (Handrick et al. 2016; Meihls et al. 2013). In this study, we extended this method to a metabolome scale, by looking for co-localized QTL across 727 mass features that are significantly different in seedling roots between two maize inbred lines, B73 and Mo17. We found that over 26% of the differentially accumulated mass features are significantly associated with one or more of six genomic hotspots scattered on four maize chromosomes (Figure 4.1; Table 4.1). This result suggests that broad diversity in specialized metabolome among maize inbred lines can be achieved through limited genetic polymorphism in their genomes. Furthermore, these genomic hotspots tend to regulate distinct segments of the maize seedling root specialized metabolome, as evident by limited overlap in mass features mapped to each of them, indicating limited shared epistasis among distinct genetic loci on each given group of specialized metabolite (Table S4.1).

Quantitative trait mapping with bi-parental populations are often limited by resolution since recombination between physically approximate loci is rare in the few generation of controlled crossing. The intermated B73-Mo17 recombinant inbred line population used in this study addressed this challenge by allowing open pollination prior to successive self-pollination to promote recombination (Lee et al. 2002). Yet, some of the metabolite QTL genomic hotspot we identified still covers over 16 Mbps of physical distance, and prioritization of a short list of candidate genes within such a large interval remains virtually impossible (Table 4.1). In this study, we hypothesized that the expression of a candidate gene should correlate with the natural variation in the metabolite it is regulating, if this regulation is achieved at transcriptional level. In two of six genomic hotspots, we observed significant elevation in mass feature-transcript correlation for genes located within the hotspot region, supporting our hypothesis, though the

same pattern is not observed in the other four hotspots (Figure 4.2). Therefore, in the limited sample size of the six genomic hotspots, we showed that the causal genetic polymorphisms associated with metabolic traits could cast their effects either through transcriptional regulation, or non-transcriptional mechanisms (*e.g.* structural variation in a metabolic enzyme, post-transcriptional activation, etc.). From a practical perspective, mass feature-transcript correlation analyses only turned out to be useful in pinpoint candidate genes within large QTL interval in one of the six genomic hotspots, since almost all the expressed genes within the genomic hotspot on the short arm of Chromosome 3 showed elevated correlation with the mass features mapped to that interval compared to genes located elsewhere (Figure 4.2). This phenomenon may be explained by co-regulation of the transcription of closely placed genes, which could be affected by epigenetic factors such as chromatin accessibility (Chen et al. 2010).

In the only case where mass feature-transcript correlation helped pinpointing a clear candidate gene, we identified a putative vesicular transport protein-encoding gene, GRMZM2G105652, as a negative regulator of 8-*O*-methylated benzoxazinoid compounds including DIM_2BOA, DIM_2BOA-Glc, and HMBOA (Figure 4.3; Table S4.1). However, results from seedling root metabolome analyses of genetic mutants were inconsistent between two mutant alleles (Figure 4.4B). Since we have not verified the potential change in target gene expression caused by transposon insertion in the 5' UTR, we cannot rule out the possibility that only one of the two (or even neither) mutant alleles has a functional effect. Unexpectedly, when we compared the two independent UFMu lines generated from the same W22 genetic background, we found clear differentiation of their seedling root metabolome, even in individuals where our target transposons were absent (Figure 4.4A). This suggests that the other *Mu* transposon insertions scattered across the genome have un-negligible effects on seedling root specialized metabolome. Therefore, the two mutant alleles of GRMZM2G105652 we obtained should be considered as in different genetic backgrounds, and any phenotypic inconsistency could also arise from interactions between our target mutation and their respective genetic backgrounds.

Sequence homology, as well as functional domain annotation of the protein encoded by our candidate gene led us to hypothesize that vesicular transport in plant cells is a negative regulator of benzoxazinoid compounds. Therefore, we tested the effect of BFA on benzoxazinoid contents in maize seedling roots. BFA is a rapid allosteric inhibitor of a vesicle coat protein

recruiter protein, and thereby inhibits the formation of vesicles traveling from rough endoplastic reticulum (ER) to Golgi apparatus in yeast, mammalian, and plant cells (Takac et al. 2011). BFA treatment on maize seedling roots can result in characteristic BFA bodies in epidermal cells in as short as 20 minutes, though this effect is reversible (Hause et al. 2006). However, prolonged BFA treatment can lead to fusion of neighboring ER and Golgi, and ultimately apoptosis (Nebenfuhr et al. 2002). In our study, we adapted a published BFA treatment protocol on maize seedling roots by prolonging the treatment period and promoting compound uptake with vacuum infiltration (Hause et al. 2006). We believed these modifications were necessary because 1) we need BFA penetration into deeper tissues than in the published study, and 2) effect of BFA treatment on benzoxazinoids would require longer time period compared to proteins and cellular compartments due to high constitutive abundance and slow turnover rate. While we did not inspect BFA-treated maize seedling roots with confocal microscopy, the metabolomic differentiation between BFA- and mock-treated samples suggests that our treatment was effective (Figure 4.5A).

Specific benzoxazinoid compounds can either be induced, suppressed, or unaffected by BFA treatment in seedling roots, suggesting that the vesicular transport process may regulate benzoxazinoid metabolism at more than one point (Figure 4.5B). Nevertheless, the total benzoxazinoid content in BFA-treated seedling roots is significantly higher than that in mock-treated samples (Figure 4.5B). This result is consistent with what we observed in one of the two UFMu mutant lines, where heterozygous individuals with supposedly lower expression of GRMZM2G105652 tend to accumulate less benzoxazinoid compounds overall, both support our original hypothesis that vesicular transport is a negative regulator of benzoxazinoid accumulation (Figure 4.4B). However, it is notable that BFA has a very broad impact on the endomembrane system in cells, and hence the specific point of regulation of vesicular transportation on benzoxazinoid metabolism requires more specific *in planta* manipulative experimentation, ideally with effective genetic mutants.

GRMZM2G105652 is predicted to encode a maize exocyst complex component 84B protein. This family of protein is highly conserved across plants, yeast, and animals, participating in a protein complex (*i.e.* exocyst) that directs post-Golgi vesicles to specific locations on the plasma membrane (Zhang et al. 2010). Hence, it participates in vesicular transport at a different part from the inhibition point of BFA. Nonetheless, both supposedly lowered expression of this

putative exocyst complex component and presence of BFA results in hyper-accumulation of total benzoxazinoids in seedling roots (Figure 4.4B; Figure 4.5B). Therefore, we hypothesize that the plant secretory pathway, which requires both ER-to-Golgi and post-Golgi vesicles, is the underlying cellular process that leads to the differential benzoxazinoid content we observed in these experiments, as well as in the two original inbred lines, B73 and Mo17. To test this hypothesis, we propose to measure both apoplastic and exuded benzoxazinoid content in the two parental lines, genetic mutants with altered secretory pathway activities, and BFA-treated seedling roots in future experiments.

Materials and Methods

Identification of metabolite QTL genomic hotspots. The same maize seedling root metabolomics dataset generated for experiments in Chapter Three was used for composite interval mapping using the same method. Since determining the significance threshold with 500 permutations for each of the 727 mass features was computationally impractical, we performed the permutation test for 10 randomly selected mass features, and took the average 95% confidence interval threshold (LOD score = 3.5) as a universal significance cutoff. For each genetic locus, we counted the number of mass features that had a LOD score above the universal significance cutoff with a custom Python script, and plotted the count of each locus along the genetic map of the maize genome constructed from the same set of genetic markers to visualize the distribution of significant metabolite QTL and identify genomic hotspots. The approximate physical locations of genetic markers flanking each metabolite QTL genomic hotspot were looked up using the online Locus Lookup tool at MaizeGDB website to determine the physical size of each genomic hotspot (Andorf et al. 2010).

Cross-omics correlation analyses. The same maize seedling root transcriptomics dataset from Chapter Three was used to calculate pairwise Pearson's correlation between each expressed transcript and mass features with co-localized QTL in a genomic hotspot. Median Pearson's correlation coefficient (R^2) of each transcript was then plotted according to its physical position in the maize genome. All statistical calculations and visualization were done in Microsoft Excel 2013 and R 3.4.0.

Amino acid sequence alignment and hidden Markov Model analyses. Predicted amino acid sequences of the two transcript variants associated with GRMZM2G105652 and their closets

Arabidopsis thaliana orthologs were downloaded from the MaizeGDB and TAIR websites. Amino acid sequence alignment of the two pairs of maize-*Arabidopsis* orthologs were performed using the MultAlign web tool (Corpet 1988). Hidden Markov Model scan of both maize amino acid sequences were performed with the hmmscan tool on the HMMER website to identify known protein domains and structural motifs (Finn et al. 2015).

Metabolomics analysis of Uniform-*Mutator* transposon insertion mutants. Seeds from two self-pollinated heterozygous mutants carrying two different *Mutator* transposon insertions in the 5' UTR of GRMZM2G105652 were planted in Turface MVP ® calcined clay (Profile Products LLC, Buffalo Grove, IL) in 3' x 3' plastic pots. Seedling roots tissues were harvested at V3 stage, and snap frozen for subsequent metabolomics analysis using the same LC-MS method and XCMS-CAMERA computational pipeline as previously described in Chapter Three. Seedling leaves tissues were also harvested for DNA extraction and polymerase chain reaction (PCR) based genotyping for the target *Mu* insertion using a pair of GRMZM2G105652-specific primers (g105652-F: 5'-CCCAACACACCCTGACAACAG-3'; g105652-R: 5'-ACCAATCCAAACCC AAGCAAAGC-3'), Terminal Inverted Repeat 6 (TIR6) primer, and TIR8 primers mix following the recommended nested PCR protocol (McCarty et al. 2013). PCA with total metabolome data and targeted quantification of specific benzoxazinoid compounds using peak areas were carried out as previously described in Chapter three. For metabolites represented by multiple mass features, peak area of the strongest mass feature (marked in bold in Table S4.1) was quantified to estimate the abundance of the metabolite.

Brefeldin A reatment and seedling root metabolomics analyses. Mo17 seeds were surface sterilized and germinated in damped germination paper as described in Chapter Two. When seedling primary roots reached approximately 5 centimeters, we submersed the roots in 50 milliliter 0.3% BFA aqueous solution (prepared from 50 μM stock solution in dimethylsulfoxide, DMSO) or equal volume of 0.3% DMSO solution as a mock treatment control, and vacuum-infiltrated for 24 hours. Seedling primary roots were then harvested and snap frozen for metabolomics analyses as described above. Each biological replicate in this comparative metabolomics experiment consisted of two pooled individuals.

References

Andorf CM, Lawrence CJ, Harper LC, Schaeffer ML, Campbell DA, Sen TZ (2010) The Locus Lookup tool at MaizeGDB: identification of genomic regions in maize by integrating sequence information with physical and genetic maps. Bioinformatics 26 (3):434-436. doi:10.1093/bioinformatics/btp556

Bollina V, Kumaraswamy GK, Kushalappa AC, Choo TM, Dion Y, Rioux S, Faubert D, Hamzehzarghani H (2010) Mass spectrometry-based metabolomics application to identify quantitative resistance-related metabolites in barley against Fusarium head blight. Mol Plant Pathol 11 (6):769-782. doi:10.1111/j.1364-3703.2010.00643.x

Chen WH, de Meaux J, Lercher MJ (2010) Co-expression of neighbouring genes in Arabidopsis: separating chromatin effects from direct interactions. Bmc Genomics 11:178. doi:10.1186/1471-2164-11-178

Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16 (22):10881-10890

Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR (2015) HMMER web server: 2015 update. Nucleic Acids Res 43 (W1):W30-38. doi:10.1093/nar/gkv397

Handrick V, Robert CAM, Ahern KR, Zhou SQ, Machado RAR, Maag D, Glauser G, Fernandez-Penny FE, Chandran JN, Rodgers-Melnik E, Schneider B, Buckler ES, Boland W, Gershenzon J, Jander G, Erb M, Kollner TG (2016) Biosynthesis of 8-O-methylated benzoxazinoid defense compounds in maize. Plant Cell 28 (7):1682-1700. doi:10.1105/tpc.16.00065

Hause G, Samaj J, Menzel D, Baluska F (2006) Fine Structural Analysis of Brefeldin A-Induced Compartment Formation After High-Pressure Freeze Fixation of Maize Root Epidermis: Compound Exocytosis Resembling Cell Plate Formation during Cytokinesis. Plant Signal Behav 1 (3):134-139

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A (2002) Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Mol Biol 48 (5-6):453-461

McCarty DR, Suzuki M, Hunter C, Collins J, Avigne WT, Koch KE (2013) Genetic and molecular analyses of UniformMu transposon insertion lines. Methods Mol Biol 1057:157-166. doi:10.1007/978-1-62703-568-2_11

Meihls LN, Handrick V, Glauser G, Barbier H, Kaur H, Haribal MM, Lipka AE, Gershenzon J, Buckler ES, Erb M, Kollner TG, Jander G (2013) Natural variation in maize aphid resistance is associated with 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one glucoside methyltransferase activity. Plant Cell 25 (6):2341-2355. doi:10.1105/tpc.113.112409

Mideros SX, Warburton ML, Jamann TM, Windham GL, Williams WP, Nelson RJ (2014) Quantitative Trait Loci Influencing Mycotoxin Contamination of Maize: Analysis by Linkage Mapping, Characterization of Near-Isogenic Lines, and Meta-Analysis. Crop Sci 54 (1):127-142. doi:10.2135/cropsci2013.04.0249

Nebenfuhr A, Ritzenthaler C, Robinson DG (2002) Brefeldin A: deciphering an enigmatic inhibitor of secretion. Plant Physiol 130 (3):1102-1108. doi:10.1104/pp.011569

Pasquet JC, Chaouch S, Macadre C, Balzergue S, Huguet S, Martin-Magniette ML, Bellvert F, Deguercy X, Thareau V, Heintz D, Saindrenan P, Dufresne M (2014) Differential gene expression and metabolomic analyses of *Brachypodium distachyon* infected by deoxynivalenol producing and non-producing strains of *Fusarium graminearum*. Bmc Genomics 15. doi:Artn 62910.1186/1471-2164-15-629

Ponts N, Pinson-Gadais L, Boutigny AL, Barreau C, Richard-Forget F (2011) Cinnamic-derived acids significantly affect *Fusarium graminearum* growth and *in vitro* synthesis of Type B trichothecenes. Phytopathology 101 (8):929-934. doi:10.1094/Phyto-09-10-0230

Richter A, Schaff C, Zhang Z, Lipka AE, Tian F, Kollner TG, Schnee C, Preiss S, Irmisch S, Jander G, Boland W, Gershenzon J, Buckler ES, Degenhardt J (2016) Characterization of Biosynthetic Pathways for the Production of the Volatile Homoterpenes DMNT and TMTT in Zea mays. Plant Cell 28 (10):2651-2665. doi:10.1105/tpc.15.00919

Santiago R, Reid LM, Arnason JT, Zhu XY, Martinez N, Malvar RA (2007) Phenolics in maize genotypes differing in susceptibility to Gibberella stalk rot (*Fusarium graminearum* Schwabe). J Agr Food Chem 55 (13):5186-5193. doi:10.1021/jf070641e

Takac T, Pechan T, Richter H, Muller J, Eck C, Bohm N, Obert B, Ren H, Niehaus K, Samaj J (2011) Proteomics on brefeldin A-treated Arabidopsis roots reveals profilin 2 as a new protein involved in the cross-talk between vesicular trafficking and the actin cytoskeleton. J Proteome Res 10 (2):488-501. doi:10.1021/pr100690f

Zarsky V, Kulich I, Fendrych M, Pecenkova T (2013) Exocyst complexes multiple functions in plant cells secretory pathways. Curr Opin Plant Biol 16 (6):726-733. doi:10.1016/j.pbi.2013.10.013

Zhang Y, Liu CM, Emons AM, Ketelaar T (2010) The plant exocyst. J Integr Plant Biol 52 (2):138-146. doi:10.1111/j.1744-7909.2010.00929.x

CHAPTER FIVE

CHEMICAL DIVERSITY AND GENETIC CONTROL OF MAIZE SPECIALIZED METABOLITES ARE REVEALED BY METABOLOME-SCALE GENOME-WIDE ASSOCIATION STUDIES[1]

Abstract

Genome-wide association analysis and metabolite-to-transcript correlation network analysis are effective means of identification of genetic control of biochemical phenotypes in plants. In this study, untargeted specialized metabolomics data are collected across the maize (*Zea mays*) Goodman-Buckler diversity panel in the tips and bases of the emerging third leaf for analysis in conjunction with publicly available high-density single nucleotide polymorphism genotyping data and transcriptomics data collected for the same tissue types across this genetic mapping population. Using the range of mass feature retention time as a proxy to chemical structural similarity, it is shown that chemical differentiation between tissue types are primarily caused by leaf tip-specific flavonoid accumulation, that tropical and temperate maize inbred lines can be differentiated based on the dominant type of benzoxazinoid compound, and that structurally similar metabolites tend to be correlated to each other in their constitutive abundance, and associated with shared genetic loci. Metabolome-scale genome-wide association studies identify large number of quantitative trait loci associated with each mass feature, including those previously found through genetic mapping with bi-parental genetic mapping populations and confirmed biochemically. In contrast to previous studies focusing on limited number of specific metabolites, most of the mass features included in this study are associated with large number of genetic loci, disregard of their prevalence or heritability in this population. This study provides a rich collection of associations between specialized metabolites and specific genetic loci, sometimes to single-gene resolution. To demonstrate the utility of this resource, we show that allelic difference in a maize citrate synthase is responsible for the accumulation of different structure isomers of hydroxycinnamic acid-hydroxycitric acid conjugates in tropical versus temperate maize inbred lines.

---

[1]Shaoqun Zhou designed the experiments, collected and analyzed all metabolomics data and the GWAS results. Kevin Ahern organized planting and tissue collection. Karl Kremling, Nonoy Bandillo, and Edward Buckler performed GWA analyses. Ying Zhang and Frank Schroeder collected and analyzed the NMR spectroscopy data. Georg Jander provided suggestions in experimental design and manuscript preparation.

Introduction

Plants produce diverse specialized metabolites that are not directly related to their central energy metabolism and structural integrity. The presence and vast diversity of these specialized metabolites are reflective of their essential functions in plant stress responses, especially in their interactions with various microbial phytopathogens and insect herbivores. For human society, plant-derived specialized metabolites have long been valuable sources of flavor, nutrition, and pharmaceutical products. More recently, advances in genetics and molecular biology have led to clarification of the complete biosynthetic pathways of many plant specialized metabolites such as glucosinolates (Halkier and Gershenzon 2006) and benzoxazinoids (Zhou et al. 2018). This knowledge has allowed researchers to manufacture some plant specialized metabolites in industrial scales, as well as to genetically improve crop species for pest and disease resistances.

Maize (*Zea mays*) is one of the most economically important staple crop species in the North America, with more than 15 billion bushels harvested in 2016, out of which more than 10% was lost to diseases (Mueller 2017). In parts of Africa, ongoing epidemics of fall armyworm (*Spodoptera frugipeda*) have devastated local maize production, and caused far-reaching socioeconomic ramifications (Stokstad 2017). These problems highlight the need for continuous genetic improvement of pest and disease resistance in the commercial maize germplasm to cope with the spatiotemporal fluctuations of biotic stresses. Even after millennia of artificial selection, maize is known for its genetic diversity at the population level (Buckler et al. 2006; Jiao et al. 2017). Similarly, different maize inbred lines are known to possess distinct profiles of specialized metabolites in various tissue types (Handrick et al. 2016; Meihls et al. 2012; Wen et al. 2014; Wen et al. 2016). Therefore, combining high-throughput metabolite profiling and existing genetic resources and genotypic data for a metabolome-scale genome-wide association studies (GWAS) would allow large-scale identification of candidate genes and loci involved in maize specialized metabolism, opening up the possibility of harnessing the natural biochemical defenses found in the broader maize germplasm for improved pest and disease resistance. Indeed, this approach has been taken to study rice seedling shoots and maize kernels, and has led to genome-wide identification of metabolic QTL in their respective study systems (Matsuda et al. 2015; Wen et al. 2014; Wen et al. 2016).

In this study, we performed liquid chromatography-mass spectrometry (LC-MS) analyses of the tips and bases of the emerging third leaf across the maize Goodman diversity panel (Flint-

Garcia et al. 2005). These two tissue types were chosen because 1) they represented distinct stages of differentiation, and 2) constitutive concentrations of specialized metabolites tended to decrease as plants age (Cambier et al. 2000; Zheng et al. 2015). The Goodman diversity panel contains 282 diverse maize inbred lines belonging to five genetic subpopulations and has been genotyped with over 29 million single nucleotide polymorphism (SNP) markers (Bukowski et al. 2015). More recently, whole transcriptome profiling of eight distinct tissue-environment combinations, including the two adopted in this study, has been done across this diversity panel (Kremling et al. 2018). Hence, the metabolomics data generated in this study can be analyzed synergistically with these existing datasets to test specific hypotheses, and to achieve a system-level understanding of specialized metabolism in maize seedling leaves.

To gain structural insights from this metabolomics dataset, we established mass feature retention time as a proxy to estimate structural relatedness among specialized metabolites, as supported by shared characteristic ultraviolet (UV) absorbance profiles associated with neighboring mass features. This allowed us to parse the diversity of maize specialized metabolites as a function of tissue types and population structure. Genome-wide association study (GWAS) with these metabolic traits allowed assessment of their genetic architecture at a global scale, and revealed shared and tissue-specific genomic hotspots that were significantly associated with numerous structurally similar metabolic traits. In addition to the genome-scale patterns, this study also provides a rich resource of high-resolution association between metabolites and genetic loci. To demonstrate the use of this resource, we confirmed the positions of known QTL affecting benzoxazinoid accumulation, and associated a group of little-studied metabolites, phenylpropanoid hydroxycitric acid esters, with a gene predicted to encode a citrate synthase-like protein. *(Functional characterization of the candidate gene in work)*

Results

Maize seedling leaf specialized metabolome is significantly different between tips and bases, but less so among genetic subpopulations.

Through high-resolution LC-MS analyses, we obtained the specialized metabolome of the leaf bases of 230 inbred lines and leaf tips of 264 inbred lines. Across this diverse maize population and two tissue types, over fifty-eight hundred mass features were detected in at least three of the samples analyzed after filtering (see Methods and Materials; Table S5.1-S5.4). To parse the

natural variation in this dataset, a principal component analysis (PCA) was performed with the top four thousand mass features (by mean intensity across all samples), and showed that tissue type explained over 30% of the variance (Figure 5.1A). Two-way analyses of variance (ANOVA) of the same dataset showed that more than 97% of all the mass features analyzed were significantly influenced by tissue type (FDR < 0.05; Figure 5.1B). In contrast, genetically defined maize population structure did not contribute to the variance significantly. Different genetic subpopulations failed to separate from each other, even when analyzed within each tissue type independently (Figure 5.1C,D), and fewer mass features were significantly different based on these subpopulations (Figure 5.1B).

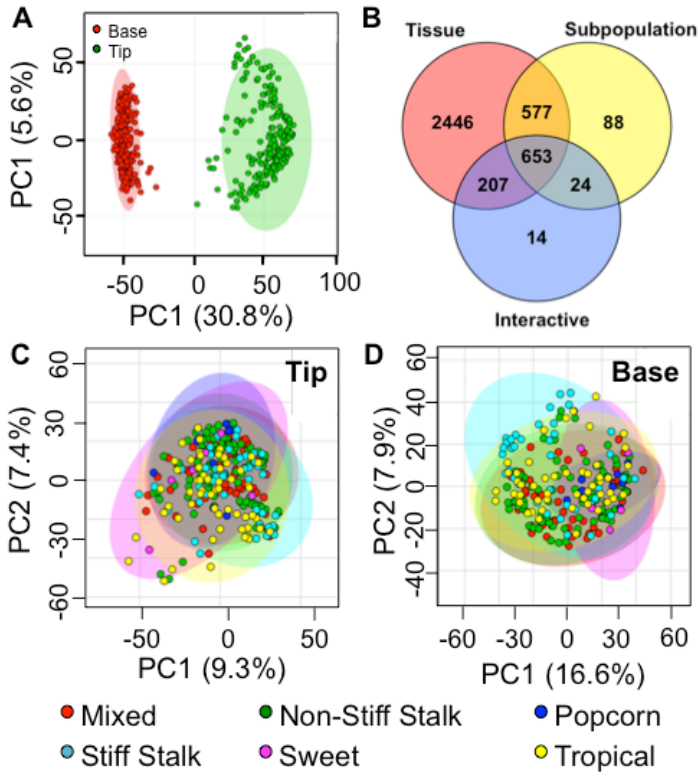


Figure 5.1. Maize leaf specialized metabolome significantly differentiates between tips and bases, but less so among genetic subpopulations. Maize seedling leaf metabolomes can be distinguished between tips and bases with a principal component analysis **(A).** Within either tissue type, genetic subpopulations cannot be differentiated based on their metabolome fingerprint **(C & D).** Consistently, many more mass features are significantly different by tissue types than by subpopulations (Two-way ANOVA, FDR < 0.05; **B**).

Metabolomic differentiation based on tissue type and genetic subpopulation are driven by different classes of specialized metabolites

On the total ultraviolet (UV) absorption chromatogram, neighboring peaks tended to have similar UV absorbance profiles. Specifically, peaks eluting between 240 and 360 seconds all had UV absorbance profiles resembling phenylpropanoids, peaks eluting between 360 and 460 seconds had typical benzoxazinoid-like UV absorbance profiles, and those eluting after 460 seconds were flavonoid-like in their UV absorbance profiles (Figure

S5.1). These observations suggested that the range of retention times of mass features could be used to assign them into one of these three major classes of maize specialized metabolites putatively.

We plotted the extent of differentiation of each mass feature based on tissue type, genetic subpopulation, or their interactive effect, as measured by the negative logarithm of p-values from two-way ANOVA, against their retention time (Figure 5.2A). These plots demonstrated that

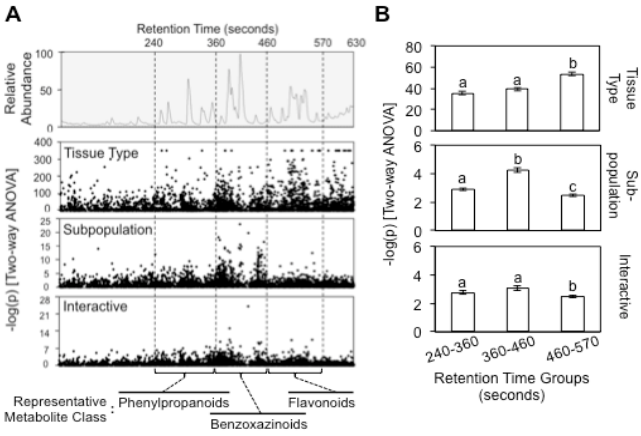


Figure 5.2. Metabolomic differentiations between tissue types and among subpopulations are driven by different classes of specialized metabolites. Each mass feature is plotted by its retention time (x-axis) and –log(p) by either tissue type, subpopulation, or their interactive effect (y-axes; Two-way ANOVA; **A**), and aligned to a sample total ion chromatogram. Average –log(p) from Two-way ANOVA by each variable are compared among three retention time ranges, corresponding to three classes of specialized metabolites, with one-way ANOVA (pairwise adjusted p-value shown between columns).

mass features from distinct ranges of the chromatogram, and hence different classes of specialized metabolites, were primarily responsible for tissue- and subpopulation-based metabolomic differentiation. Specifically, mass features that were significantly different between leaf tips and bases tended to fall in the range of flavonoids, whereas those under significant influence from the maize subpopulation or its interaction with tissue type were almost exclusively found among the benzoxazinoids (Figure 5.2A). These visual patterns were confirmed with statistical comparison of the extent of differentiation between retention time groups (Figure 5.2B). These observations suggested that 1) flavonoids were significantly different between the leaf tip and leaf base samples, and that 2) inbred lines belonging to different genetic subpopulations were significantly different in their benzoxazinoid contents. In support of the first hypothesis, all major mass signals co-eluting with flavonoid-like UV absorption peaks were completely absent in leaf base samples, and only found in the more developed leaf tips. Furthermore, the only two genes encoding chalcone synthases, the enzyme catalyzing the first committing step in flavonoid biosynthesis, expressed in these two tissue types (based on B73 reference genome ver. 4, Annotation 5b+; Jiao et al., 2017) were both expressed significantly higher in the leaf tips across this diversity panel (Figure S5.2). To test the second

hypothesis, we picked out the mass features representing the most abundant benzoxazinoid compounds in maize seedlings, 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA) and its methylated glucoside derivative, 2-(2-hydroxy-4,7-dimethoxy-1,4-benzoxazin-3-one)-β-D-glucopyranose (HDMBOA-Glc). DIMBOA-Glc was significantly depleted in the tropical inbred lines. Instead, the tropical inbred lines showed significantly higher abundance of HDMBOA-Glc, and presumably its highly unstable and undetectable aglucone (Figure S5.3). This pattern is consistent with results from a previous study of the maize nested association mapping parental lines, where the authors identified a deactivating transposon insertion in the DIMBOA-Glc methylation enzyme BX12 in temperate, but not tropical maize inbred lines (Meihls et al. 2012).

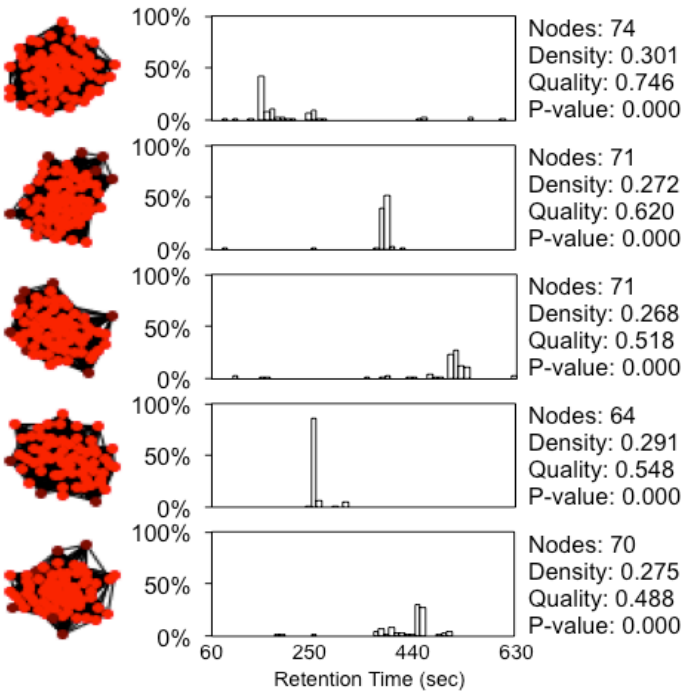


Figure 5.3. Mass features in the same correlation network tend to have similar retention times. Distributions of retention times of mass features within the same correlation network are plotted in 10 seconds incremental bins and are aligned to the topographical presentations of the networks. Density, quality, and p-value of each network is calculated by the graph-clustering algorithm ClusterOne (Nepusz et al., 2012). Only results for the 5 biggest clusters are presented here. Results for all 21 significant clusters ($p < 0.05$; one-sided Mann-Whitney U test) are shown in Table S5.2.

Structurally related metabolites tend to be co-regulated.

Since structurally related metabolites usually arise from shared metabolic pathways, we hypothesize that these metabolites may be co-regulated in plants, and hence their natural variation should correlate with each other across the population. To test this hypothesis, we constructed mutual rank-based correlation networks with the metabolomic dataset using three different exponential decay functions, and detected overlapping correlative clusters using the ClusterONE algorithm (Nepusz et al. 2012; Wisecaver et al. 2017). With the leaf tip data, significant overlapping clusters were detected including a total of 824 mass features, with a median cluster size of 54 nodes. In support of our hypothesis, mass features

in 15 of the 21 significant clusters were closely placed together based on their retention time (Figure 5.3; Table S5.5).

Maize pan-specialized metabolome is skewed towards rare metabolites.

Our dataset provides a great opportunity to examine the diversity of specialized metabolites in

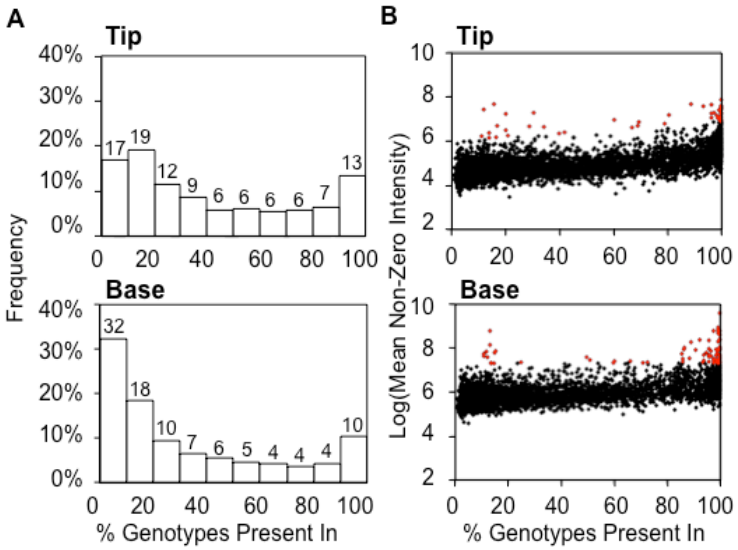


Figure 5.4. Mass feature occurrence rates are bi-modally distributed and are positively correlated with their average non-zero intensity. (A) Distribution of mass feature occurrence rate in either tissue type is plotted in 10% incremental bins. Occurrence rate in percentage (%) of each bin is labeled on top of each column. (B) Each mass feature in either tissue type is plotted by its occurrence rate (x-axis) and the log of average non-zero intensity scale (y-axis). Significant positive linear correlations between the two variables are found in both tissue types. Mass features that fall outside of the 95% confidence interval of the overall linear correlation patterns are marker in red.

maize. In both tissue types, there was a bi-modal frequency distribution of mass feature occurrence rate, measured by percent of genotypes where a mass feature was detected. Specifically, approximate 15% of mass features in either tissue type were present in more than 90% of all the genotypes, whereas more than 63% of mass features are found in less than half of the genotypes examined (Figure 5.4A). Rare mass features could result from rare metabolites or MS noises, while the later tend to be lower in signal intensity compared to mass features resulted from true metabolites. The mean non-zero intensity of each mass feature showed significant positive correlations ($R^2 > 0.96$) with its occurrence rate in both tissue types (Figure 5.4B), suggesting that rare mass features indeed tended to be lower in abundance. Yet, the slope of either linear regression was smaller than one on the logarithmic scale (0.972 in tips and 0.908 in bases), meaning that on average, a mass feature detected in 10% of all the genotypes would only be less than ten-fold lower in intensity than a ubiquitous mass feature. In contrast, mass features of any occurrence rate could vary by a hundred fold (Figure 5.4B). Therefore, the rare mass features were true metabolites that were present in only a subset of tested maize inbred lines.

Genome-wide association studies with known metabolites reveal both known and novel genetic loci.

The existing genotype dataset for the Goodman Diversity Panel (Bukowski et al. 2015) makes it possible to perform GWAS with each mass feature as an independent trait to understand its genetic architecture. Given the large number of traits to be analyzed, a recently developed fast GWAS pipeline using a general linear model optimized was adopted for rapid recursive GWAS (Kremling et al. 2018). Two benzoxazinoid compounds, 2-(2,4-dihydroxy-7,8-dimethoxy-1,4-benzoxazin-3-one)-β-d-glucopyranose (DIM_2BOA-Glc) and HDMBOA-Glc, in leaf tips were chosen as positive controls. Biosynthetic genes of these two compounds have been identified through genetic mapping with bi-parental recombinant inbred lines populations (Handrick et al. 2016; Meihls et al. 2012).

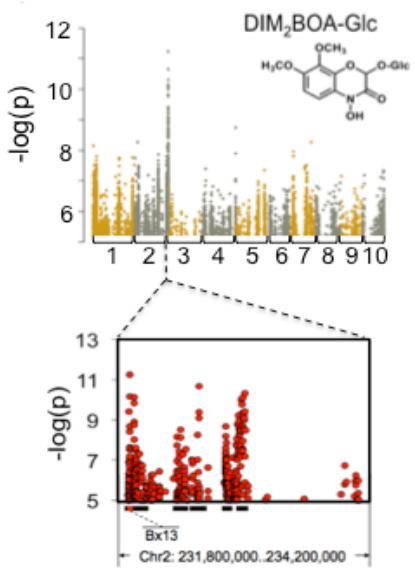


Figure 5.5. DIM_2BOA-Glc is significantly associated with genetic markers within and adjacent to its immediate biosynthetic gene. Natural variation in the abundance of DIM_2BOA-Glc was mapped by GWAS. Each SNP marker is plotted by its physical location in the maize genome (x-axis) and level of association with DIM_2BOA-Glc abundance (y-axes). SNP markers on adjacent chromosomes (labeled on the bottom) are shown in different colors. Only SNP markers with –log(p) > 5 are plotted. Local LD blocks around the most highly associated markers, calculated from the same SNP dataset, are indicated by black bars at the bottom of the plots, and known benzoxazinoid biosynthetic genes are highlighted in red.

GWAS with both metabolites identified their respective biosynthetic genes (*Bx10-12* for HDMBOA-Glc and *Bx13* for DIM_2BOA-Glc), with the most significantly associated SNPs being in linkage disequilibrium (LD) with the respective genes (Figure 5.5 & 5.6A). Interestingly, in addition to the SNP markers in LD with the known biosynthetic genes, GWAS also identified SNP markers associated with the metabolites of interest in adjacent LD blocks (Figure 5.5 & 5.6A). Further away, a locus on Chromosome 9 was strongly associated with natural variation in HDMBOA-Glc abundance (Figure 5.6A), and appeared to act additively with the *BX10-12* locus on Chromosome 1 (Figure 5.6B). The top 4 most significant SNPs (p = 0.000) were located within a single 25 kilo base pairs (kb) LD block, which contained the 3' parts of GRMZM2G108309 (Figure 5.6A). This gene model encoded a putative protein phosphatase 2C family protein. 3' RNAseq data obtained from comparable tissues showed that the expression of GRMZM2G108309 was significantly different between the inbred lines carrying either allele at

the LD block (Figure 5.6C). Furthermore, inbred lines with high GRMZM2G108309 expression accumulated significantly higher amount of HDMBOA-Glc than those with low expression (Figure 5.6D). This preliminary evidence suggested that this gene model could indeed be a novel regulator of HDMBOA-Glc content in the tips of maize seedling leaves.

Genetic architecture of specialized metabolites tends to be complex, and varies independently from heritability and occurrence rate.

Since GWAS with HDMBOA-Glc and DIM_2BOA-Glc successfully identified known genetic loci, we believe the fast GWAS pipeline and our LC-MS dataset are capable of correctly pinpointing genetic loci regulating specialized metabolites on a metabolome scale. Prior to this computation-intensive analysis, the metabolomics dataset was further filtered by the broad sense heritability ($H^2 \geq 0.2$), and rate of occurrence (detected in ≥ 10% of all genotypes examined). The augmented experimental design allowed estimation of broad sense heritability by calculating

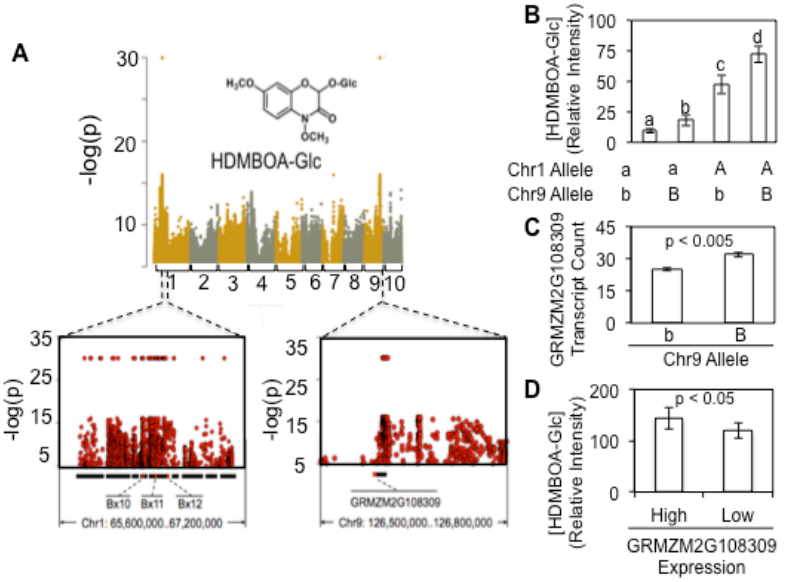


Figure 5.6. Genome-wide association analysis with HDMBOA-Glc identifies a putative novel regulator in addition to known biosynthetic genes. Natural variation in the abundance of HDMBOA-Glc was mapped by GWAS. Each SNP marker is plotted by its physical location in the maize genome (x-axis) and level of association with HDMBOA-Glc abundance (y-axes). SNP markers perfectly associated with the phenotype (*i.e.* p = 0) was rounded down to 30 on the y-axes for graphical representation. SNP markers on adjacent chromosomes (labeled on the bottom) are shown in different colors. Only SNP markers with –log(p) > 5 are plotted. Local LD blocks around the most highly associated markers, calculated from the same SNP dataset, are indicated by black bars at the bottom of the plots, and known benzoxazinoid biosynthetic genes are highlighted in red (A). Additive effect on HDMBOA-Glc abundance of the two loci on chromosome 1 and chromosome 9 (B), effect of allelic segregation on the candidate gene expression (C), and the effect of candidate gene expression on HDMBOA-Glc abundance (D) were plotted. Error bars = s.e.

the variance of mass features measured in replicated B73 control samples, which were planted in each flat, as the environmental variance. Mass features not detected in B73 but present in other inbred lines were retained, even though their broad sense heritability could not be estimated. 1,319 mass features from the leaf bases and 2,672 mass features from the leaf tips remained after these filtering processes, and GWAS was performed with each of these metabolites using the 29 million SNP dataset (Bukowski et al. 2015).

Previous genetic mapping analyses with metabolic traits in rice and maize have often identified small numbers of large effect-size genetic loci (Chen et al. 2014; Handrick et al. 2016; Matsuda et al. 2015; Meihls et al. 2012; Wen et al. 2014). To investigate whether these observations represent the rule or the exception in the genetic architecture of metabolic traits, the top 10 most strongly associated SNP markers for each mass feature were collected, and the number of 10 Kbps LD blocks they mapped to were counted. This analysis showed that less than 9% of all the mass features analyzed have their top 10 most strongly associated SNP markers located in three or less 10 Kbps LD blocks (Figure 5.7A). This indicated that metabolic traits, unlike previously assumed, tended to have complex genetic architecture, under control of numerous interacting genetic loci. No consistent trend in genetic architecture complexity, as measured by number of mapped loci, was observed with increasing heritability or occurrence rate (Figure 5.7B,C). The only consistent pattern was that the most commonly

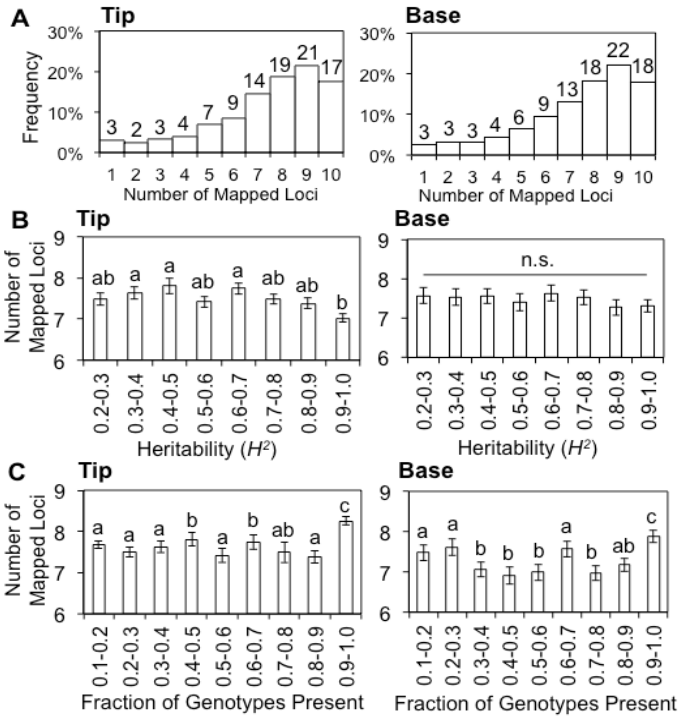


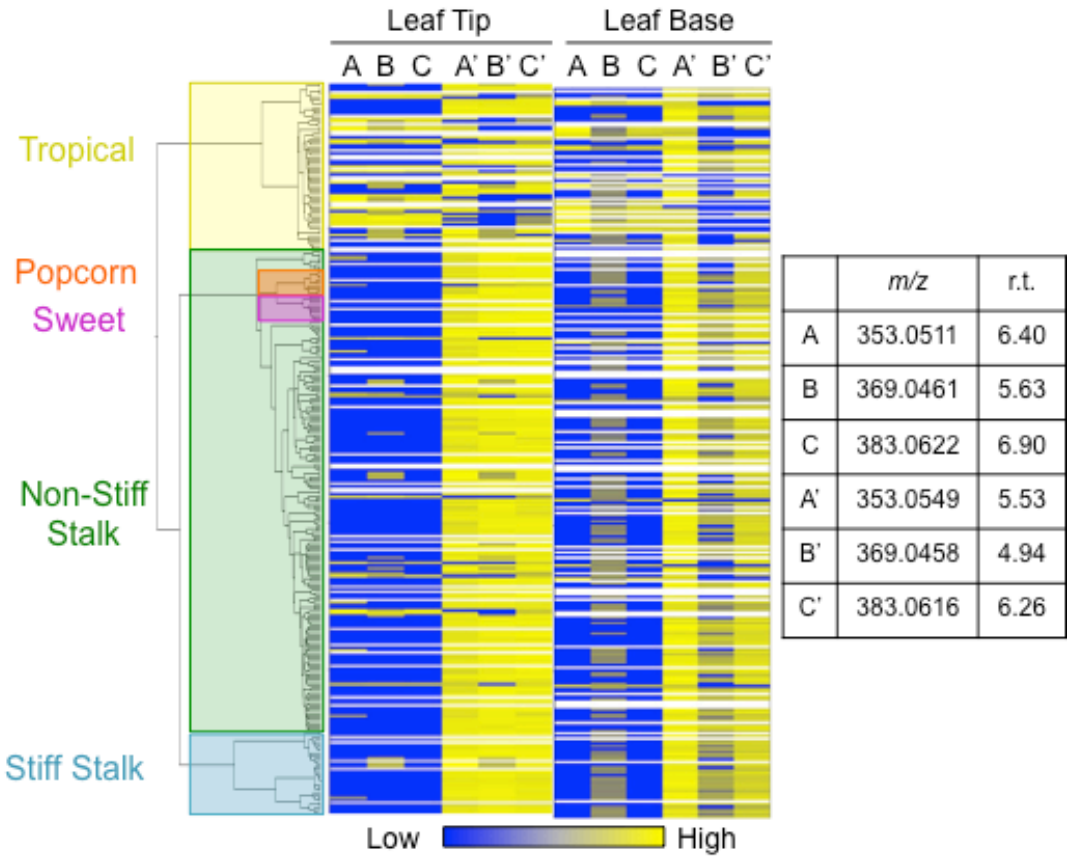
Figure 5.7. Metabolic traits tend to have complex genetic architecture irrespective of their heritability or occurrence rate. (A) Distribution of mass features in either tissue type is plotted by number of 10 kbp LD blocks it is strongly associated with. This measurement is then compared across different heritability level (B) and occurrence rate (C) by one-way ANOVA and Tukey HSD. Groups significantly different from each other ($p < 0.05$) are denoted with different letters on their respective columns. Error bars = standard errors.

found mass features (occurrence rate > 90%) tended to map to significantly more loci than mass features that were less prevalent in the population (Figure 5.7C).

Different phenylpropanoid hydroxycitric acid ester isomers found in distinct maize subpopulations are associated with a putative citrate synthase family protein.

The ultimate goal of most genetic mapping projects is to identify loci that are significantly associated with specific metabolites of interest, pinpoint candidate genes, and experimentally test for a causal relationship. Here, we demonstrate the utility of maize metabolite GWAS by associating structural polymorphism in an uncharacterized biosynthetic gene with the abundance of a little-studied class of specialized metabolites.

One of the most outstanding patterns that manifested itself in the previous analyses of



	m/z	r.t.
A	353.0511	6.40
B	369.0461	5.63
C	383.0622	6.90
A'	353.0549	5.53
B'	369.0458	4.94
C'	383.0616	6.26

Figure 5.8. Three pairs of phenylpropanoid-containing structural isomers have complementary distribution across the maize diversity panel. A phylogenetic tree of the 282 maize inbred lines included in the GWAS panel is constructed with the distance matrix calculated from 66 thousand SNP markers. The estimated concentration of three pairs phenylpropanoid-containing structural isomers are shown in a color scale for each maize inbred line ordered by their phylogenetic relationships. These metabolites are represented by mass features detected in negative ionization mode (*m/z*) and retention time in minutes (r.t.). Each monophyletic groups are assigned to a genetic subpopulation as defined in Flint-Garcia et al., 2005 by the predominant group assignment for the individuals within that clade.

specialized metabolite diversity was that there were clear outliners to the overall positive correlation between the occurrence rate and mean non-zero intensity of mass features (Figure 4B). While the majority of these outliners were concentrated on the high occurrence rate range, where the linear correlative relationship was capped by maximal occurrence rate, in both leaf tips and bases, a group of high intensity mass features were only detected in around 20% of all the examined genotypes. Among these outliers, five were found in both leaf tips and bases (Table S5.6). Two of these five high-abundance, low-occurrence mass features appeared to be fragments of 2-methyl-3,1-benzoxazin-4-one (MBOA), a spontaneous non-enzymatic degradation product of diverse benzoxazinoid compounds found in high abundance in juvenile maize tissues (Zhou et al. 2018). These compounds probably resulted from sporadic post-extraction degradation in some samples even though the extracts were stored at -20°C before LC-MS analyses. The other three mass features likely represented three different metabolites since they eluted at distinct retention times. Interestingly, these three metabolites all had characteristic phenylpropanoid-like UV absorbance profiles, and had two common daughter ions around *m/z* = 189.004 and *m/z* = 127.003 under negative electron spray ionization (Figure S5.4A). This suggested that these three metabolites were conjugates of the same compound to different phenylpropanoid moieties. The mass differences between the three metabolites suggested that they could be interconverted through a hydroxylation and a methylation reaction.

In maize inbred lines where these metabolites were not detected, at least one other peak was present at that specific *m/z* channel at a different retention time. These peaks also co-eluted with the same probable daughter ions, and UV absorption peaks with phenylpropanoid-like absorbance profiles. Therefore, these peaks were probably structural isomers of the three high-abundance, low-occurrence metabolites, but with higher polarity and occurrence rate (Figure S5.4B). To examine how these pairs of structural isomers were distributed across the maize inbred line population, a phylogeny of the population was estimated using a 66 thousand SNP dataset (see Materials and Methods), and the abundance of their representative mass features across this phylogenetic tree was plotted. In both tissue types, the rare isomers tended to co-occur with each other, and were over-represented among the tropical inbred lines (Figure 5.8). Furthermore, the two groups of isomers tended to be mutually exclusive. It was noteworthy though, that these general trends were not perfect, especially in the case of the isomers with *m/z* = 369.046, both of which were sporadically distributed across the population in the leaf bases

without necessarily co-occurring with the other metabolites. The metabolism of these pairs of phenylpropanoid-containing isomers could be under developmental regulation as demonstrated by the different phylogenetic patterns in leaf tips and leaf bases.

Two-dimensional nuclear magnetic resonance spectroscopy data revealed that these three compounds are ester conjugates of coumaric acid, caffeic acid, and ferulic acid, respectively, with hydroxycitric acid (Table S5.7). Although these molecules have been previously isolated and identified from maize, they have not been evaluated for potential bioactivity (Ozawa et al. 1977). GWAS showed that all three phenylpropanoid-hydroxycitric acid esters were strongly associated with a locus on Chromosome 4. Specifically, the most strongly associated SNP markers for all three phenylpropanoid hydroxycitric acid esters all locate within a 10 Kbps LD block (Figure S5.5). In the B73 reference genome, this LD block contained a single gene model, GRMZM2G063909, which was annotated as an ortholog of *Arabidopsis thaliana* and *Oryza sativa* citrate synthase family genes (Figure 5.9). Expression of this gene was not significantly different between maize inbred lines accumulating different phenylpropanoid-hydroxycitric acid esters (data from Kremling et al., 2018; Figure S5.6). We therefore hypothesize that there is structural polymorphism in this gene model between maize inbred lines of contrasting metabolic phenotypes.

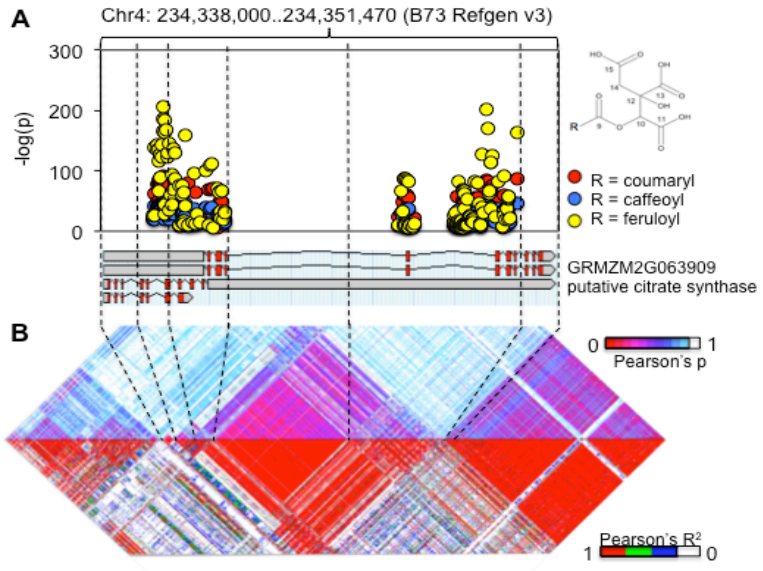


Figure 5.9. Phenylpropanoid hydroxycitric acid esters are associated with a predicted citrate synthase-like gene. SNP markers most strongly associated with the phenypropanoid hydroxycitric acid esters are plotted by their physical location in the maize genome (x-axis) and level of association with the metabolites (y-axis), and overlaid on the predicted transcripts of GRMZM2G063909 located at the same locus (A). Pairwise correlation coefficients between SNP markers around the candidate gene are calculated to demonstrate that the significantly associated SNP markers are not in linkage disequillibrium with any adjacent gene model (B).

Structurally related metabolites tend to be regulated by shared genetic loci.

In addition to identifying candidate genes significantly associated with individual metabolites of interest,

the metabolome-scale GWAS results can be used to find genetic loci with disproportional influence on the overall maize specialized metabolome. To that end, the top 10 or 50 most significant SNP markers for each of the 4,859 mass features were identified and their distribution across the maize genome was plotted. This identified a number of loci, to which many mass features have mapped in either leaf tip or leaf base tissues (Figure 10 A&C).

In both tissue types, three loci on Chromosome 1, 4, and 10 showed higher-than-background number of mass feature GWAS hits. Specifically, the genomic hotspot on Chromosome 1 was localized to a 110 Kbps region containing the *BX11* and *BX12* genes, two paralogous *O*-methyltransferases that catalyzed the biosynthesis HDMBOA-Glc (Meihls et al. 2012). Indeed, mass features mapped to this locus included those related to HDMBOA-Glc, DIMBOA, and other benzoxazinoid compounds. Interestingly, many mass features not associated with known benzoxazinoid compounds were also mapped to this locus, suggesting that it could also regulate other specialized metabolites. The genomic hotspot on Chromosome 4, which contained the most mass features GWAS hits in both tissue types, was divided

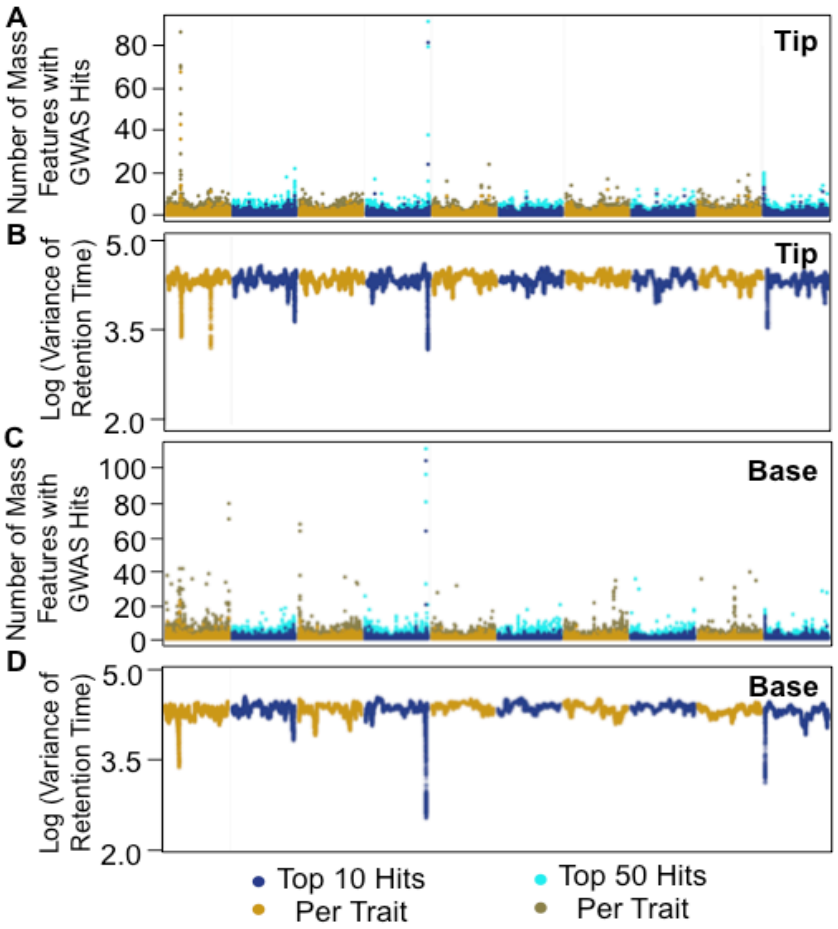


Figure 5.10. Metabolite GWAS hotspots tend to be associated with neighboring mass features. (A,C) The number of mass features at least one of their top 10 or top 50 most strongly associated SNP marker located in each 10 kbps LD block is plotted for either tissue type. (B,D) Variance in the retention time of 100 mass features with closely placed GWAS hits are calculated and mapped across the genome based on the physical location of the first SNP hits. Results of neighboring chromosomes are shown in different colors, and results based on different top SNP threshold are indicated by different color brightness.

between two 10 Kbps LD blocks that contained the two putative citrate synthase family proteins discussed above (GRMZM2G063909 and GRMZM2G064023). Finally, the less prominent hotspot on Chromosome 10, which influenced a dozen mass features in either tissue type, spanned a 30 Kbps region. In the reference B73 genome, this region contains seven transposon genes and a low confidence gene model. The prevalence of transposon genes in this region in the B73 genome suggests that there is likely whole-gene level presence/absence variation caused by transposon activity among the diverse maize inbred lines.

In addition to the three hotspots that shared between both tissue types, there were also a handful of genomic hotspots specific to either tissue type. For example, 9 mass features had at least one of their 10 most associated SNP markers located within a 20 Kbps region on Chromosome 3. This region contained a single gene model, GRMZM2G143723, analogous to a rice C2H2 zinc finger protein. These tissue type-specific genomic hotspots were indicative of development-dependent regulation of specialized metabolism in maize seedling leaves.

Among these genomic hotspots, two contained confirmed or likely biosynthetic genes. In the case of the Chromosome 1 hotspot containing *BX11* and *BX12*, we observed that most of the mass features mapped to this locus represented benzoxazinoid metabolites. This example led us to hypothesize that these genomic hotspots contained one or more genes that regulate multiple metabolites derived from the same biosynthetic pathway, which were hence structurally related. To test this hypothesis, the mass features were ordered by where their most strongly associated SNP markers were located in the maize genome, and the variance in retention time of one hundred adjacent mass features was calculated with a sliding window. Across the maize genome, there was a stable background level of the variance in retention time of mass features mapped to close physical positions in both tissue types. However, there were clear dips in this measurement below the background level at some loci. When results from this analysis were aligned to the previous counting of mass features per locus, there was a clear co-localization of dips in retention time variance with the genomic QTL hotspots (Figure 10 B&D), suggesting that structurally related metabolites tended to be co-regulated by the same genetic loci. This pattern was true for all three of the genomic hotspots shared by both tissue types, but not necessarily valid all the time. For example, the significant dip in variance of retention time on Chromosome 1 downstream of the genomic hotspot containing *BX11* and *BX12* did not correspond to any

increase in the number of mass features mapped to that locus, whereas mass features mapped to the leaf base-specific hotspot on Chromosome 3 did not have very similar retention times.

Discussion

Technological advances in mass spectrometry and accumulating high-density genotypic data are allowing increasing number of metabolome-scale quantitative genetics studies of metabolic traits. Existing studies of this type range from primary metabolites of nutritional interest to known specialized metabolites in both model plants and economically relevant crop species (Chan et al. 2010; Chan et al. 2011; Chen et al. 2014; Matsuda et al. 2015; Riedelsheimer et al. 2012; Wen et al. 2014). However, unlike transcriptomic data, where each transcript can be functionally annotated to some extent based on sequence homology and structural features, most mass features from untargeted metabolomics datasets represent unknown metabolites, and the mass spectrometry data provide incomplete information about their potential structures. In addition to large scale co-elution tests with known compounds, one of the current methods to address this issue is to construct molecular networks based on shared tandem mass spectrometry (MS/MS) fragments, which are indicative of structural similarity (Matsuda et al. 2015; Nguyen et al. 2013).

In the current study, we propose that mass feature retention time also can be used as proxy of structural relatedness among metabolites in absence of MS/MS fragmentation data, as supported by shared UV absorbance profiles of major mass features within certain retention time ranges (Supplemental Figure 2). Though the resolution and accuracy of structural-relatedness estimate by retention time is not as good as the prediction by shared MS/MS fragments, it is more universally applicable, especially for low intensity mass features and metabolites that do not yield detectable fragments. Using this proxy, we found evidence that 1) metabolic differentiation between tissue types and genotypic subpopulations were driven by different classes of specialized metabolites (Figure 2), 2) structurally related metabolites tended to be co-regulated (Figure 3; Table S5.2), and 3) structurally related metabolites tended to be associated with shared genomic loci (Figure 9).

Our metabolomics dataset allowed us to assess the chemical diversity of maize specialized metabolome in two tissue types across a diverse maize population. The metabolomes of both tested tissue types demonstrated bi-modal distributions, with a relatively small core

component and a large number of rare metabolites (Figure 4A). Unfortunately, this analysis had not been done in published population-scale metabolomics study in maize and rice for comparison. Instead, we compared these distributions to the published maize pan-transcriptome, and found that the distribution of our metabolomics data were much more left-skewed, probably reflective of the more diverse nature of specialized metabolites compared to the transcriptomes, which contained large number of housekeeping genes involved in primary metabolism and cellular structures (Hirsch et al. 2014).

We demonstrated the quality of the high-resolution genetic mapping results from the metabolite GWA analyses performed in this study by confirming known biosynthetic loci and identifying novel candidate genes associated with two known benzoxazinoid compounds (Figure 5&6), as well as a class of previously little-studied maize metabolites, phenylpropanoid hydroxycitric acid esters (Figure 9). Furthermore, we provided a metabolome-scale evaluation of the complexity of genetic architecture of metabolic traits. Unexpectedly, only a small number of metabolic traits seemed to have a simple genetic architecture, as measured by the number of loci significantly associated with them, and we observed no consistent correlative relationship between genetic architecture complexity and heritability or occurrence rate of metabolic traits (Figure 7). This observation echoed a recent report where widespread genetic epistasis between maize kernel metabolic QTL was indicated (Wen et al. 2016). Another potential explanation for this observed pattern could be that the same metabolic trait is regulated by different genetic loci in different subsets of the maize population. This hypothesis could also explain the significantly higher number of the mapped loci associated with the most ubiquitous mass features (Figure 7C).

In addition to interpreting GWAS results with specific known and previously unknown metabolites, we also identified shared and tissue-specific metabolite QTL hotspots by compiling metabolome-scale GWAS results (Figure 10). This non-uniform distribution of significant GWAS hits was comparable to results from a published rice metabolite GWAS (Chen et al. 2014). These loci provide promising candidates for breeding and genetic engineering of specialized metabolism and biochemical defense in maize as they have disproportional influence on the metabolome.

Materials and Methods

Plant growth and tissue collection. All maize seeds were originally obtained from the Maize Genetics Cooperation Stock Center (Urbana Champaign, Illinois). To ensure comparability of our metabolomics data with previous published transcriptomics data collected in the same tissue types, the exact same seed stocks were used and the growth conditions was replicated in the same greenhouse space (Kremling et al., 2018). Eight seeds of each maize genotype were planted, and the entire diversity panel was fitted into twenty-six 96-cell flats. To control for micro-environmental variation, eight B73 seeds were included in each flat, and all flats were randomized daily. When the third leaf had visibly emerged from the whorl, two centimeters of tissue from the tips and bases were collected. For leaf base tissues, seedlings were cut down at soil line, and unrolled to expose the base. For each maize inbred line, tissues from two seedlings were pooled, weighed, and snap frozen in liquid nitrogen for metabolite extraction.

Metabolomics analyses and data preprocessing. Frozen seedling leaf tissues were extracted with 200 microliters of acidified 50% methanol, and analyzed on a Sigma Supleco reverse phase C18 column on a Dionex 3000 Ultimate UPLC-diode array detector system coupled to Thermo Q Exactive mass spectrometer. The two mobile phase solvents were water (Solvent A) and acetonitrile (Solvent B), both acidified with 0.1% formic acid. The mobile phase gradient ran from 95% Solvent A at 0 minutes to 100% Solvent B at 10.5 minutes with curvature of 2 to optimize compound separation, while reducing the runtime of each individual analysis to accommodate our large sample size. Each extract was analyzed with both positive and negative modes of electron spray ionization separately. Metabolite quantification was estimated with signal intensity acquired through the XCMS-CAMERA mass scan data processing pipeline (Benton et al. 2010; Kuhl et al. 2012; Tautenhahn et al. 2008). To maximize the detectable chemical diversity, the minimal samples threshold for keeping a mass feature was set at three at the grouping step of the XMCS processing.

Mass features detected by the XCMS-CAMERA pipeline were filtered based on their retention times (60-630 seconds) and exact masses ($m/z < 0.5$ at first decimal point), and peaks annotated as naturally occurring isotopes were removed. Mass feature quantification was then corrected by tissue fresh weight and the total ion concentration of each sample to account for technical variation.

Chemical diversity analyses. Measurement of each mass feature across the diversity panel was log-transformed for multivariate analyses. A small positive number was added to each zero value prior to log-transformation. This dataset was uploaded to the MetaboAnalyst 3.0 online tool platform for principal component analysis and Two-way ANOVA. The mass feature list was further filtered by interquartile range, Pareto scaled, and only the top four thousand mass features (by mean intensity across all samples) were used for these analyses. Each maize inbred line was assigned to a genetic subpopulation as defined in Flint-Garcia et al. (2005). All other statistical analyses and data visualization were carried out in R and Microsoft Excel 2013.

Correlative network analyses. The same metabolomic datasets were used to calculate pairwise Pearson correlation matrices, and then mutual rank matrices for the two tissue types separately. Pairwise mutual rank indices were converted to edge weights by an exponential decay function with $\lambda = 25$as previously described (Wisecaver et al. 2017). For each conversion, edges with weight lower than 0.01 were filtered out. These edge lists were imported into Cytoscape v 3.4.0 (Shannon et al. 2003) and overlapping clusters were detected with the ClusterONE app (Nepusz et al. 2012).

Genome-wide association study with metabolic traits. Signal intensity of each mass feature across the population was log-transformed. Box-cox transformation was skipped as it distorted the distribution of the rare mass features with lots of zero values. Mass features were filtered based on estimated broad sense heritability and rate of occurrence as described in the Result section, and the remaining 3,991 mass features were analyzed with the fast GWAS pipeline. To reduce data storage to a realistic level, only SNPs with $-\log(p) \geq 5$ for each mass feature were recorded.

To survey the genetic architectures of metabolic traits, and investigate their relationship with trait heritability and occurrence rate at a metabolomic scale, the top 10 most strongly associated SNP markers for each metabolic trait to three different fixed size LD blocks, namely 10 kb, 60 kb, and 360 Kbps were mapped. As expected, more metabolic traits have their top GWAS SNP hits located within fewer number of LD blocks as

the estimated LD size increases, but the overall shape of distribution was not affected (Supplemental Figure 7). The same LD-block assigning process was used to generate the overview of GWAS hits distribution across the maize genome, by counting the numbers of mass features mapped to each LD block, and plotting them according to the physical location of the

LD blocks in the maize genome. Similarly, the locations of metabolite QTL genomic hotspots are consistent across different window sizes of LD (Supplemental Figure 8). Finally, GWAS hits were ordered based on their physical location in the maize genome, and the log variance of mass feature retention time of a hundred adjacent hits was calculated using a sliding window algorithm.

Local LD estimation and inbred lines phylogenetic reconstruction. SNP marker data across the same GWAS diversity panel around the most strongly associated SNP makers for each trait were downloaded from Cyverse Discovery Environment under the following directory (iplant/home/shared/panzea/hapmap3/hmp321), and used to estimate local LD with the pairwise correlation with sliding window algorithm implemented in TASSEL 5.2.40 (Bradbury et al. 2007). A much small SNP dataset from Samayoa et al. (2015), with filtering for maximal missing data (<80%), maximal heterozygosity level (<50%), and minimal minor allele frequency (>30%) to estimate the phylogenetic relationship among the maize inbred lines included in this study. Approximately 66 thousand SNP markers were retained after the filtering process and they were used to calculate a pairwise distance matrix with TASSEL 5.2.40. This distance matrix was then used to construct a phylogenetic tree using a hierarchical clustering algorithm with the Ward method implemented by the hclust function in R.

References

Andorf CM, Cannon EK, Portwood JL, Gardiner JM, Harper LC, Schaeffer ML, Braun BL, Campbell DA, Vinnakota AG, Sribalusu VV, Huerta M, Cho KT, Wimalanathan K, Richter JD, Mauch ED, Rao BS, Birkett SM, Sen TZ, Lawrence-Dill CJ (2016) MaizeGDB update: new tools, data and interface for the maize model organism database. Nucleic Acids Res 44 (D1):D1195-D1201. doi:10.1093/nar/gkv1007

Benton HP, Want EJ, Ebbels TMD (2010) Correction of mass calibration gaps in liquid chromatography-mass spectrometry metabolomics data. Bioinformatics 26 (19):2488-2489. doi:10.1093/bioinformatics/btq441

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23 (19):2633-2635. doi:10.1093/bioinformatics/btm308

Buckler ES, Gaut BS, McMullen MD (2006) Molecular and functional diversity of maize. Curr Opin Plant Biol 9 (2):172-176. doi:10.1016/j.pbi.2006.01.013

Bukowski R, Guo X, Lu Y, Zou C, He B, Rong Z, Wang B, Xu D, Yang B, Xie C, Fan L, Gao S, Xu X, Zhang G, Li Y, Jiao Y, Doebley J, Ross-Ibarra J, Buffalo V, Buckler ES, Xu Y, Lai J, Ware D, Sun Q (2015) Construction of the third generation Zea mays haplotype map. bioRxiv. doi:10.1101/026963

Cambier V, Hance T, de Hoffmann E (2000) Variation of DIMBOA and related compounds content in relation to the age and plant organ in maize. Phytochemistry 53 (2):223-229. doi:Doi 10.1016/S0031-9422(99)00498-7

Chan EK, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ (2011) Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in Arabidopsis thaliana. PLoS Biol 9 (8):e1001125. doi:10.1371/journal.pbio.1001125

Chan EK, Rowe HC, Hansen BG, Kliebenstein DJ (2010) The complex genetic architecture of the metabolome. PLoS Genet 6 (11):e1001198. doi:10.1371/journal.pgen.1001198

Chen W, Gao Y, Xie W, Gong L, Lu K, Wang W, Li Y, Liu X, Zhang H, Dong H, Zhang W, Zhang L, Yu S, Wang G, Lian X, Luo J (2014) Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. Nat Genet 46 (7):714-721. doi:10.1038/ng.3007

The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45 (D1):D158-D169. doi:10.1093/nar/gkw1099

Cook CE, Bergman MT, Cochrane G, Apweiler R, Birney E (2017) The European Bioinformatics Institute in 2017: data coordination and integration. Nucleic Acids Res. doi:10.1093/nar/gkx1154

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44 (6):1054-1064. doi:10.1111/j.1365-313X.2005.02591.x

Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Annu Rev Plant Biol 57:303-333. doi:10.1146/annurev.arplant.57.032905.105228

Handrick V, Robert CAM, Ahern KR, Zhou SQ, Machado RAR, Maag D, Glauser G, Fernandez-Penny FE, Chandran JN, Rodgers-Melnik E, Schneider B, Buckler ES, Boland W, Gershenzon J, Jander G, Erb M, Kollner TG (2016) Biosynthesis of 8-O-methylated benzoxazinoid defense compounds in maize. Plant Cell 28 (7):1682-1700. doi:10.1105/tpc.16.00065

Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Penagaricano F, Lindquist E, Pedraza MA, Barry K, de Leon N, Kaeppler SM, Buell CR (2014) Insights into the maize pan-genome and pan-transcriptome. Plant Cell 26 (1):121-135. doi:10.1105/tpc.113.119982

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS, Guill K, Regulski M, Kumari S, Olson A, Gent J, Schneider KL, Wolfgruber TK, May MR, Springer NM, Antoniou E, McCombie WR, Presting GG, McMullen M, Ross-Ibarra J, Dawe RK, Hastie A, Rank DR, Ware D (2017) Improved maize reference genome with single-molecule technologies. Nature 546 (7659):524-527. doi:10.1038/nature22971

Kremling KAG, Chen SY, Su MH, Lepak NK, Romay MC, Swarts KL, Lu F, Lorant A, Bradbury PJ, Buckler ES (2018) Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. Nature 555 (7697):520-523. doi:10.1038/nature25966

Kuhl C, Tautenhahn R, Bottcher C, Larson TR, Neumann S (2012) CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Anal Chem 84 (1):283-289. doi:10.1021/ac202450g

Lai JS, Dey N, Kim CS, Bharti AK, Rudd S, Mayer KFX, Larkins BA, Becraft P, Messing J (2004) Characterization of the maize endosperm transcriptome and its comparison to the rice genome. Genome Research 14 (10a):1932-1937. doi:10.1101/gr.2780504

Matsuda F, Nakabayashi R, Yang Z, Okazaki Y, Yonemaru J, Ebana K, Yano M, Saito K (2015) Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. Plant J 81 (1):13-23. doi:10.1111/tpj.12681

Meihls LN, Kaur H, Jander G (2012) Natural variation in maize defense against insect herbivores. Cold Spring Harbor symposia on quantitative biology 77:269-283. doi:10.1101/sqb.2012.77.014662

Mueller D (2017) Corn disease loss estimates from the United States and Ontario, Canada - 2016. Purdue University Extension, Retrieved from http://cropprotectionnetwork.org/crop-loss-estimates/corn-disease-loss-estimates-2016/.

Nepusz T, Yu H, Paccanaro A (2012) Detecting overlapping protein complexes in protein-protein interaction networks. Nat Methods 9 (5):471-472. doi:10.1038/nmeth.1938

Nguyen DD, Wu CH, Moree WJ, Lamsa A, Medema MH, Zhao X, Gavilan RG, Aparicio M, Atencio L, Jackson C, Ballesteros J, Sanchez J, Watrous JD, Phelan VV, van de Wiel C, Kersten RD, Mehnaz S, De Mot R, Shank EA, Charusanti P, Nagarajan H, Duggan BM, Moore BS, Bandeira N, Palsson BO, Pogliano K, Gutierrez M, Dorrestein PC (2013) MS/MS networking guided analysis of molecule and gene cluster families. Proc Natl Acad Sci U S A 110 (28):E2611-2620. doi:10.1073/pnas.1303471110

Ozawa T, Nishikiori T, Takino Y (1977) Three New Substituted Cinnamoyl Hydroxycitric Acids from Corn Plant. Agricultural and Biological Chemistry 41 (2):359-367. doi:10.1080/00021369.1977.10862499

Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proc Natl Acad Sci U S A 109 (23):8872-8877. doi:10.1073/pnas.1120813109

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst Biol 61 (3):539-542. doi:10.1093/sysbio/sys029

Samayoa LF, Malvar RA, Olukolu BA, Holland JB, Butron A (2015) Genome-wide association study reveals a set of genes associated with resistance to the Mediterranean corn borer (Sesamia nonagrioides L.) in a maize diversity panel. BMC Plant Biol 15:35. doi:10.1186/s12870-014-0403-3

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13 (11):2498-2504. doi:10.1101/gr.1239303

Solovyev V, Kosarev P, Seledsov I, Vorobyev D (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol 7. doi:ARTN S10 DOI 10.1186/gb-2006-7-s1-s10

Stokstad E (2017) New crop pest takes Africa at lightning speed. Science 356 (6337):473-474. doi:10.1126/science.356.6337.473

Tautenhahn R, Bottcher C, Neumann S (2008) Highly sensitive feature detection for high resolution LC/MS. Bmc Bioinformatics 9. doi:Artn 504 10.1186/1471-2105-9-504

Wen W, Li D, Li X, Gao Y, Li W, Li H, Liu J, Liu H, Chen W, Luo J, Yan J (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nat Commun 5:3438. doi:10.1038/ncomms4438

Wen W, Liu H, Zhou Y, Jin M, Yang N, Li D, Luo J, Xiao Y, Pan Q, Tohge T, Fernie AR, Yan J (2016) Combining Quantitative Genetics Approaches with Regulatory Network Analysis to Dissect the Complex Metabolism of the Maize Kernel. Plant Physiol 170 (1):136-146. doi:10.1104/pp.15.01444

Wisecaver JH, Borowsky AT, Tzin V, Jander G, Kliebenstein DJ, Rokas A (2017) A Global Co-Expression Network Approach for Connecting Genes to Specialized Metabolic Pathways in Plants. Plant Cell 29 (5):944-959. doi:10.1105/tpc.17.00009

Zheng LL, McMullen MD, Bauer E, Schon CC, Gierl A, Frey M (2015) Prolonged expression of the BX1 signature enzyme is associated with a recombination hotspot in the benzoxazinoid gene cluster in Zea mays. J Exp Bot 66 (13):3917-3930. doi:10.1093/jxb/erv192

Zhou S, Richter A, Jander G (2018) Beyond defense: Multiple functions of benzoxazinoids in maize metabolism. Plant Cell Physiol. doi:10.1093/pcp/pcy064

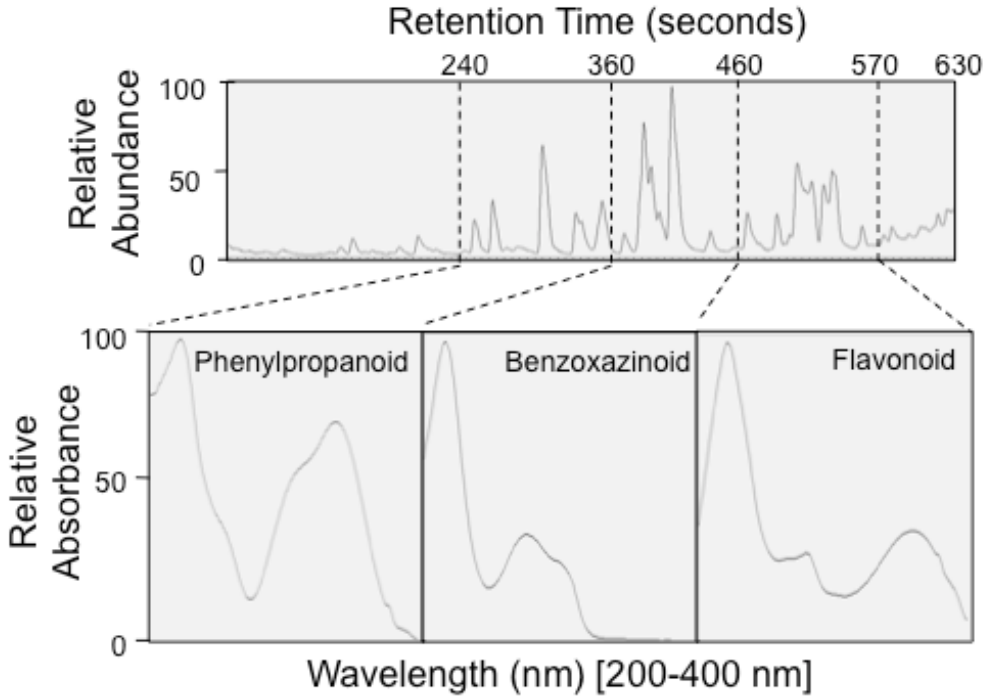


Figure S5.1. Major peaks from distinct ranges of the chromatogram share characteristic UV absorbance profiles. The UV absorbance profile of each peak was constructed by a photodiode array detector. He boundary of each retention time range is determined by the last peak showing the same characteristic UV absorbance profile. There is no detectable overlap between neighboring retention time ranges.

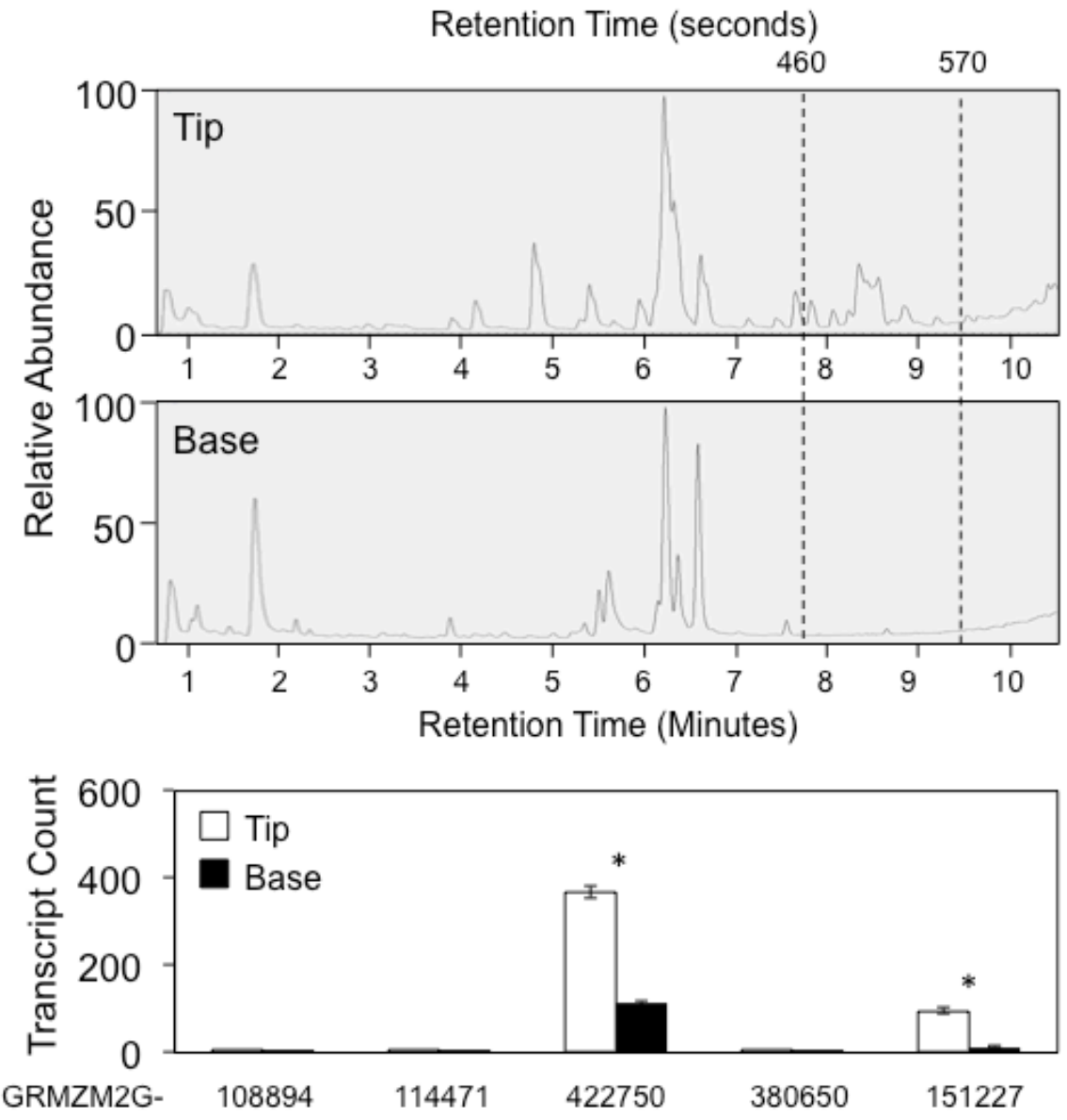


Figure S5.2. Flavonoids are absent and chalcone synthases expressions are low in seedling leaf base tissues. Sample UV absorbance chromatograms of the seedling leaf tip and base of the same genotype are shown to demonstrate the lack of peaks in the flavonoid time range (460-570 seconds). Average expressions of five chalcone synthase-encoding gene models in the B73 reference genome v4 (Annotation 5b+) across the Goodman diversity panel are compared between these two tissue types with Student's *t*-tests (*FDR < 0.05). Error bars = standard errors. Expression data obtained from Kremling et al., 2018.

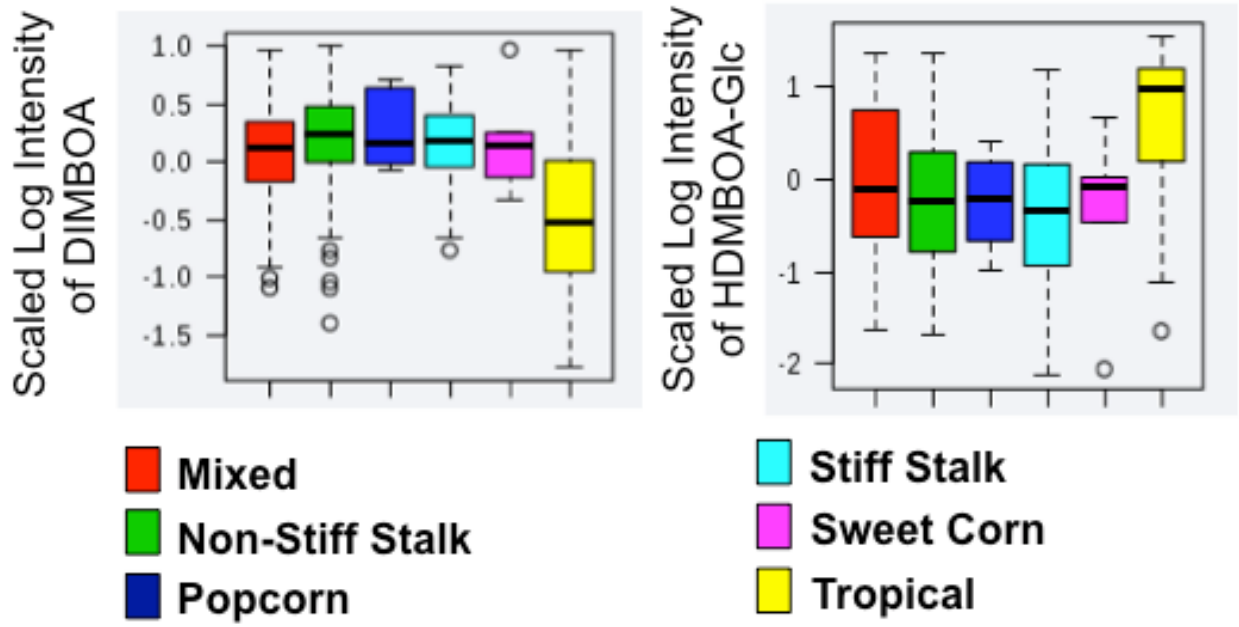


Figure S5.3. Tropical and temperate maize lines accumulate different benzoxazinoid compounds. Maize inbred lines are assigned to genetic subpopulations defined in Flint-Garcia et al., 2005. The average of each group is denoted by the central think black line, the ranges by the whiskers, and the quartiles by the boxes.

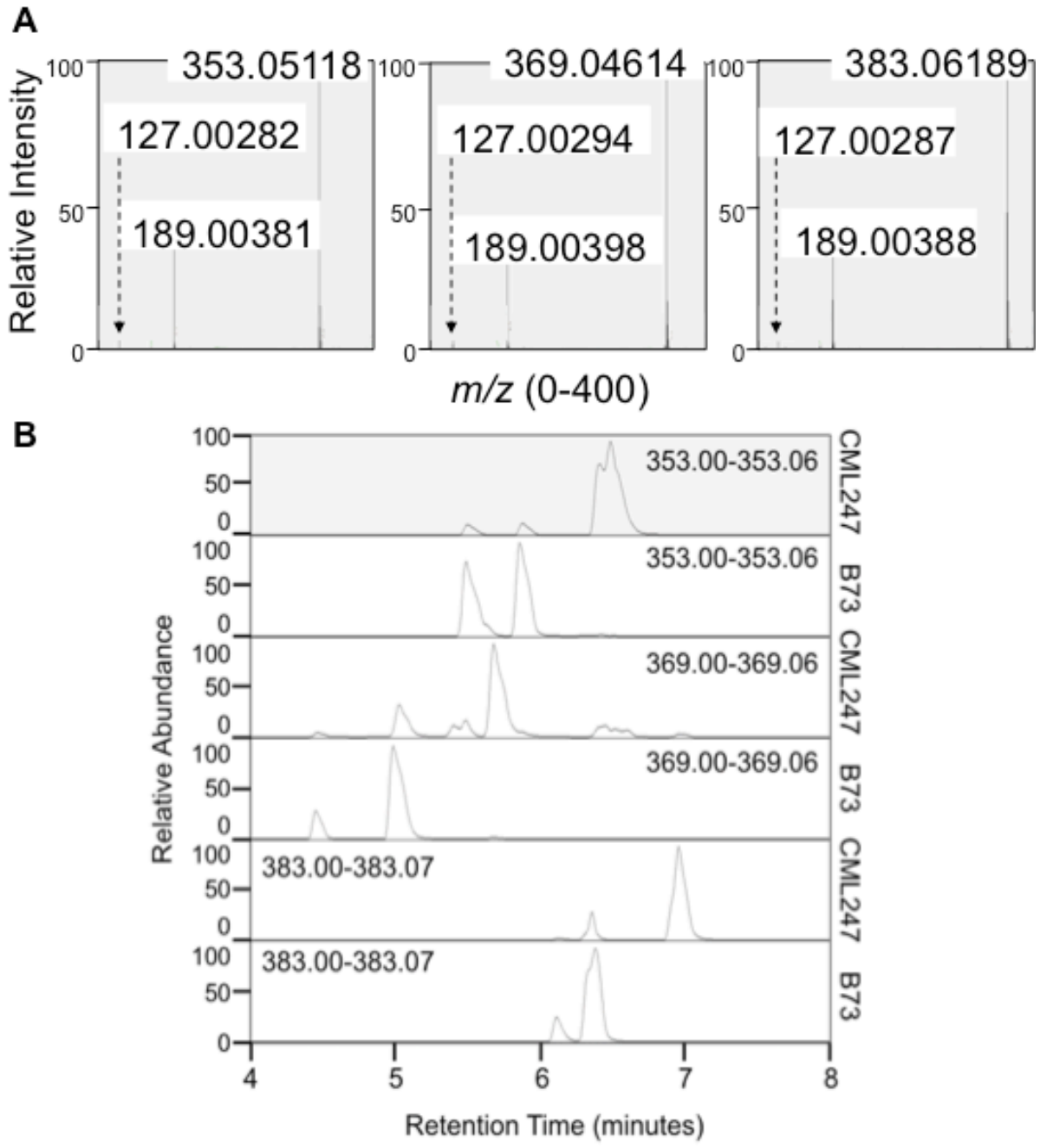


Figure S5.4. Phenylpropanoid-like mass features co-elute with common daughter ions. (A) Mass spectrum scans of three mass features co-eluting with phenylpropanoid-like UV absorbance peaks are shown. The parental ions and the two shared daughter ions are labeled with their exact *m/z* measurement. (B) In two different maize inbred lines, the predominant peaks at each specific *m/z* range eluting at different retention times likely represent different structural isomers of the same compound.

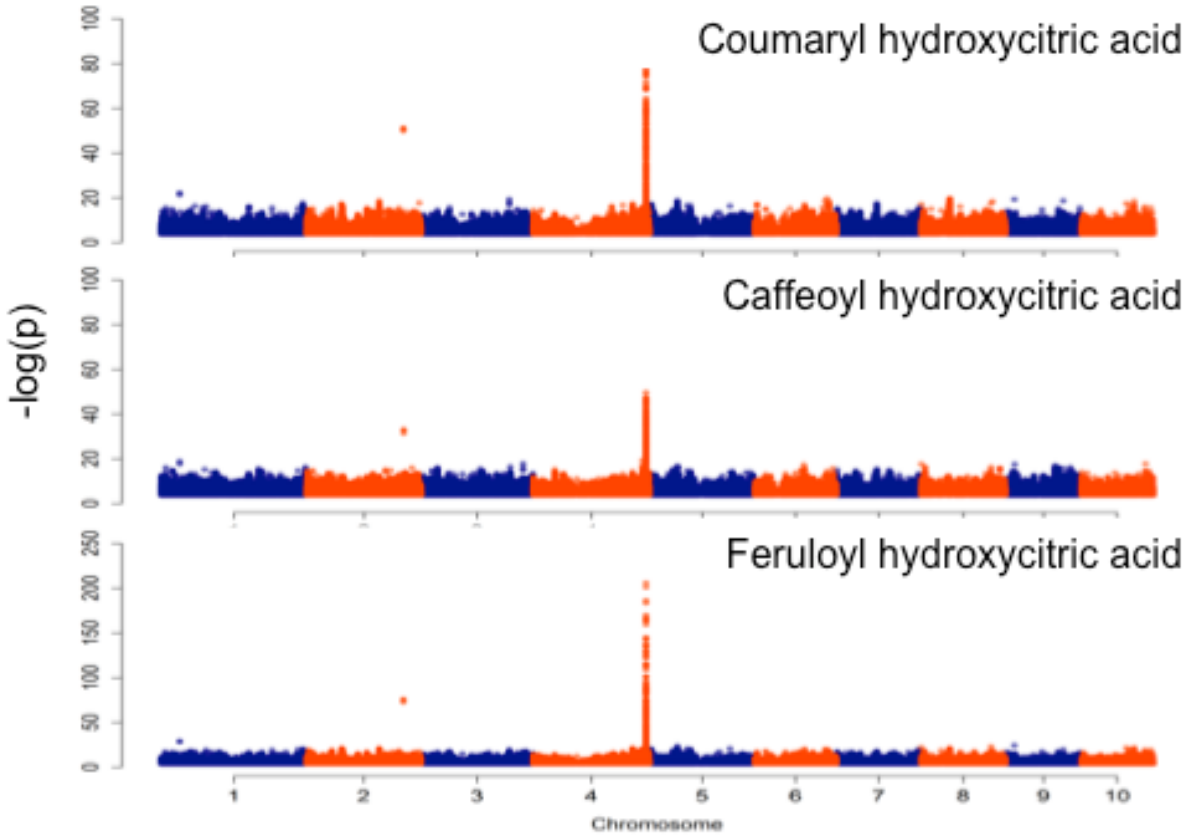


Figure S5.5. Phenylpropanoid hydroxycitric acid esters are strongly associated with a shared locus on chromosome 4. SNP markers are plotted by their physical location in the maize genome (x-axis) and level of association with phenylpropanoid hydroxycitric acid esters (y-axes). Only makers with at least –log(p) > 5 are plotted. Markers on neighboring chromosomes are shown in different colors.

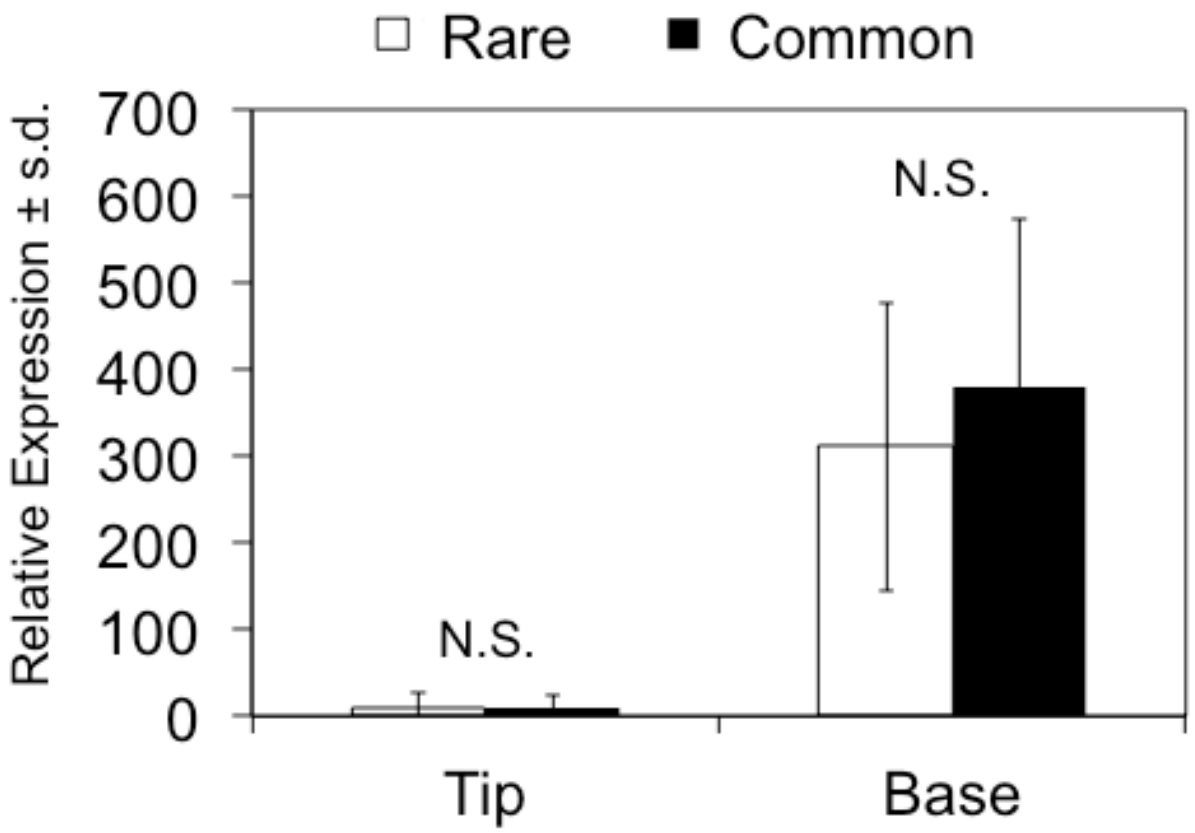


Figure S5.6. Expression of the citrate synthase family gene is not significantly different between maize inbred lines accumulating different phenylpropanoid hydroxycitric acid isomers. GRMZM2G063909 expression in the two tissue types under study were obtained from Kremling et al., (2018), and compared between maize inbred lines accumulating different phenylpropanoid hydroxycitric acid ester isomers. No significant difference (N.S.) in expression is found in either tissue type ($p > 0,05$; Student's *t*-test).

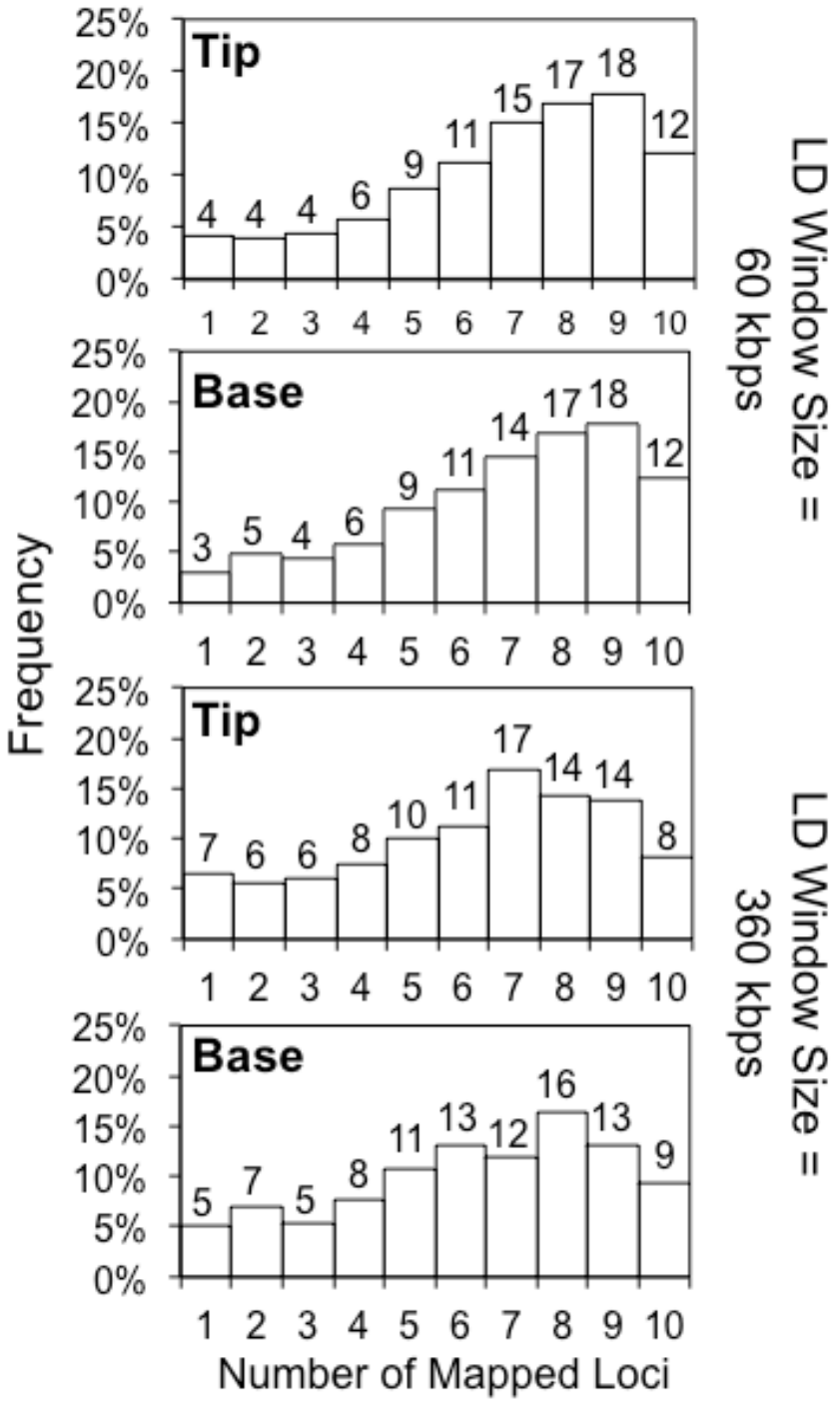


Figure S5.7. Frequency distributions of genetic architecture complexity of metabolic traits are consistent across different LD window sizes. Distribution of mass features in either tissue type is plotted by number of 60 Kbps or 360 Kbps LD blocks they are strongly associated with.

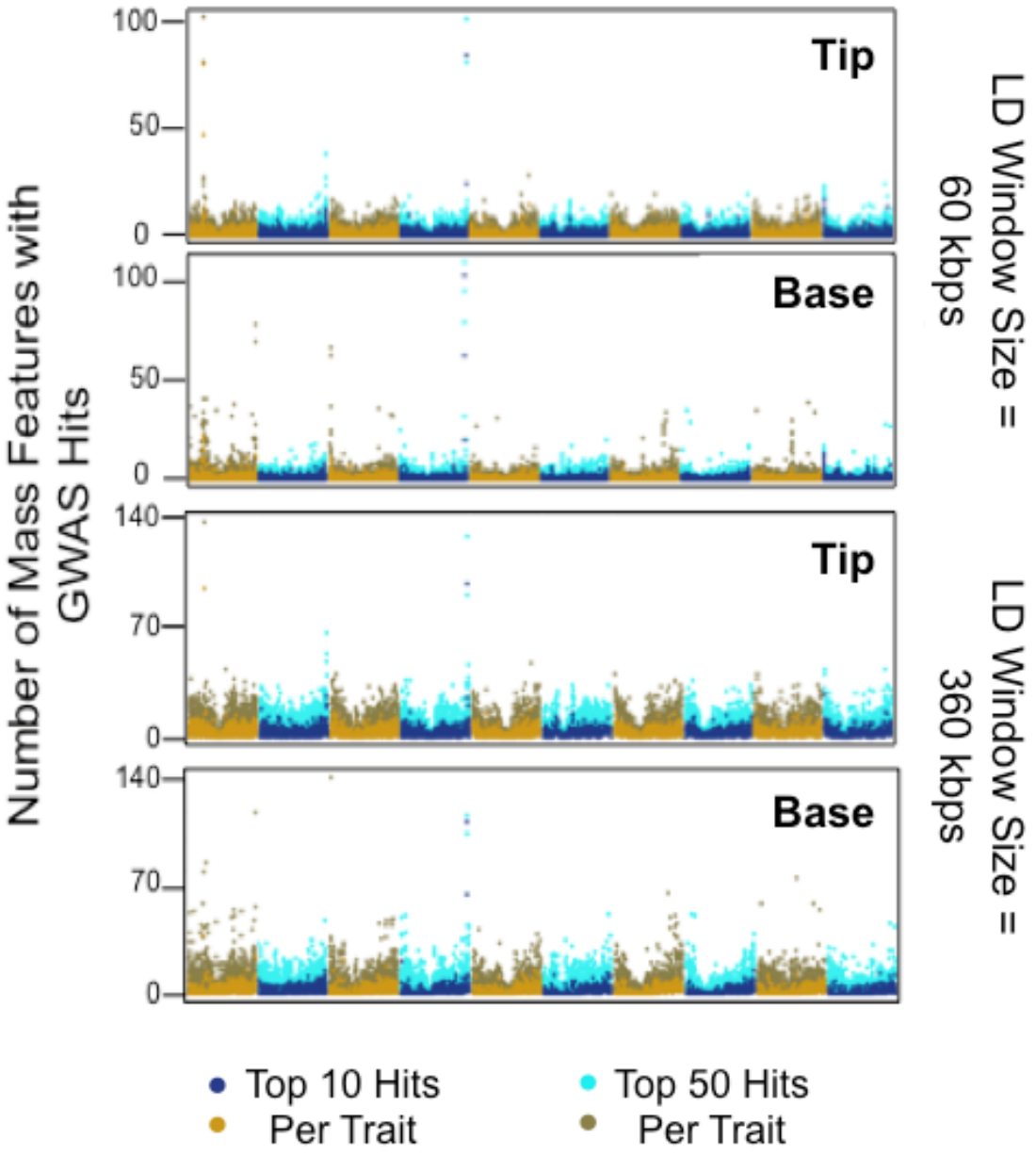


Figure S5.8. Presence and locations of metabolite QTL hotspots are consistent across different LD window sizes. The number of mass features with at least one of their top 10 or top 50 most strongly associated SNP marker located in each 60 kbps or 360 kbps LD block is plotted for either tissue type.

CHAPTER SIX

TOWARDS QUANTITATIVE MODELING OF PLANT METABOLISM: CONCLUSIONS AND FUTURE OUTLOOK

Since 2010, large-scale parallel metabolite quantification and high-density genetic marker generation through high-throughput sequencing have enabled metabolome-scale genetic mapping and cross-omics correlation analyses in various model and crop species (Chan et al. 2011; Chan et al. 2010; Riedelsheimer et al. 2012). These studies have shown their power in extending our knowledge on plant specialized metabolism and its regulation by providing a large number of clear candidate genes (Chen et al. 2014; Jin et al. 2017; Wen et al. 2014). With the maturation of various genome-editing tools in crop species, large-scale validation of these candidate genes, especially the metabolic enzyme-encoding genes with clear impact on metabolic phenotypes, should become routine in the near future. Yet, as demonstrated in two studies in this thesis (Chapter Three and Four), natural variation in specialized metabolites can also be attributed to polymorphism in physiological and cellular regulatory processes. This pattern would probably be even more common among primary metabolic traits, where complete presence/absence polymorphism in a biosynthetic gene would be intolerable. In these cases, chemical hypotheses-driven inquiry would no longer be helpful since the exact chemical structure of a metabolite does not give any hint about the putative functional annotation of a candidate gene. This problem can be, and has been addressed by increasing genetic mapping resolution with denser genetic markers and more diverse GWAS panels. However, such massive genetic resources would probably be still limited to maize and other economically relevant crop species in the near future.

In addition to the quantitative genetics approach, co-expression network analyses and gene cluster scan based on genomic features have also been useful tools in gene discovery (Frey et al. 1995; Nutzmann et al. 2016; Wisecaver et al. 2017). However, these tools are also biased towards tightly co-regulated biosynthetic pathways. In this thesis, metabolome-transcriptome correlation analyses have been used to identify non-enzyme-encoding genes regulating specialized metabolites both in a small bi-parental recombinant inbred line population and a larger GWAS population. Yet, both attempts were met with limited success at best: the regulatory function of the putative vesicular transport protein on benzoxazinoids was only

partially supported by follow-up experiments, whereas abundance of known metabolites and the expression of their biosynthetic genes failed to correlate with each other.

Among the many possible explanations for why this approach is not as productive as one may hope I believe two factors have heavy influences and can be improved with limited additional efforts. Firstly, while transcriptional regulation of metabolites is probably widespread, the relationship between each specific pair of gene and metabolite is not necessarily linear. Rather, I believe this relationship is probably more commonly sigmoidal, with both the gene expression and the metabolite abundance demonstrating bi-modal distributions. As an example, neither smilaside A nor smiglaside C content in the seedling roots among the inter-mated B73-Mo17 recombinant inbred lines show significant correlation with the expression of *ZmEIN2* in our study (data not shown). The candidate gene only manifested itself when I segregated the population based on the bi-modal distribution of the biochemical phenotype, and compared the two phenotypic classes. Secondly, whether we consider a linear or a sigmoidal relationship between each pair of gene and metabolite, it is highly unlikely that a metabolite is only affected by a single gene. Rather, as demonstrated in our GWAS results, metabolic traits also tend to have complex genetic architecture (Figure 5.6). Both quantitative genetics and cross-omics correlation approaches are biased towards genetic loci or genes of large effect size, and lack the power to detect minor contributors, even though they may be of less interests from a practical perspective. Intriguingly, both of these pitfalls can be simultaneously addressed by adopting an artificial neural network framework. Under this framework, a metabolite (output) is regulated by multiple transcripts (input), and the relationships between input and output are non-linear. Therefore, one of the immediate next steps should be developing and training an artificial neural network using both data presented in this thesis and other published cross-omics datasets.

Other than potential algorithmic problems in analyzing the metabolome-transcriptome dataset, one should also question the assumption of transcriptional regulation of plant specialized metabolism. From the phenylpropanoid hydroxycitric acid ester example presented in Chapter Five, we see that natural variation in specialized metabolites can also arise from non-transcriptional variation, such as structural polymorphism in a biosynthetic enzyme. Other mechanisms such post-transcription processing, translation, and post-translational modification could also hypothetically contribute to this variation. To-date, there is no relative estimate of

how often each of these categories is the leading cause of variation in metabolic traits, though it should be calculable from various metabolome-scale GWAS results.

Finally, analyzing the large quantitative dataset generated from these studies have led me to question the qualitative models that are often presented as the final summaries of plant physiology and molecular biology studies. For example, in our final model of how ethylene and ethylene sensitivity regulate smilaside A and smiglaside C metabolism, we had to incorporate a semi-quantitative element (*i.e.* ethylene sensitivity promote stronger smilaside A to smiglaside C conversion than the step before) in order to remain consistent with most of our experimental results. Stemming from this experience, I wonder how the stimulus-response curve would look like even for textbook examples of plant responses to biotic stresses. How would plants respond to different level of pathogenic infection, insect herbivore attack, or mere jasmonic acid spray? From classical plant development studies, it should have become clear that even the same stimulus at different magnitude would incite different responses, and that plant metabolism cannot be simplified as qualitative activation and suppression between interacting components. As we slowly get accustomed to the time when data acquisition routinely outpace data analysis, the accumulation of massive dataset should hopefully push for the paradigm shift towards quantitative modeling of plant metabolism.

Reference

Chan EK, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ (2011) Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. PLoS Biol 9 (8):e1001125. doi:10.1371/journal.pbio.1001125

Chan EK, Rowe HC, Hansen BG, Kliebenstein DJ (2010) The complex genetic architecture of the metabolome. PLoS Genet 6 (11):e1001198. doi:10.1371/journal.pgen.1001198

Chen W, Gao Y, Xie W, Gong L, Lu K, Wang W, Li Y, Liu X, Zhang H, Dong H, Zhang W, Zhang L, Yu S, Wang G, Lian X, Luo J (2014) Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. Nat Genet 46 (7):714-721. doi:10.1038/ng.3007

Frey M, Kliem R, Saedler H, Gierl A (1995) Expression of a cytochrome P450 gene family in maize. Mol Gen Genet 246 (1):100-109

Jin M, Zhang X, Zhao M, Deng M, Du Y, Zhou Y, Wang S, Tohge T, Fernie AR, Willmitzer L, Brotman Y, Yan J, Wen W (2017) Integrated genomics-based mapping reveals the genetics underlying maize flavonoid biosynthesis. BMC Plant Biol 17 (1):17. doi:10.1186/s12870-017-0972-z

Nutzmann HW, Huang A, Osbourn A (2016) Plant metabolic clusters - from genetics to genomics. New Phytol 211 (3):771-789. doi:10.1111/nph.13981

Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proc Natl Acad Sci U S A 109 (23):8872-8877. doi:10.1073/pnas.1120813109

Wen W, Li D, Li X, Gao Y, Li W, Li H, Liu J, Liu H, Chen W, Luo J, Yan J (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nat Commun 5:3438. doi:10.1038/ncomms4438

Wisecaver JH, Borowsky AT, Tzin V, Jander G, Kliebenstein DJ, Rokas A (2017) A Global Co-Expression Network Approach for Connecting Genes to Specialized Metabolic Pathways in Plants. Plant Cell 29 (5):944-959. doi:10.1105/tpc.17.00009